Aus dem Deutschen Krebsforschungszentrum
Geschäftsführender Direktor: Prof. Dr. Michael Baumann
Forschungsschwerpunkt: Infektion, Entzündung und Krebs
Sprecher Prof. Dr. Ralf Bartenschlager

Identification of single-stranded DNAs from beef products, their replication in human cells and their relationship with breast cancer

Inauguraldissertation

zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)

an der

Medizinischen Fakultät Heidelberg

der

Ruprecht-Karls-Universität

Vorgelegt von
Konstantina Falida

aus Patras, Griechenland
2018

Dekan: Herr Prof.Dr. med. Wolfang Herzog

Doktorvater: Herr Prof. Dr. med. Dr. h.c. mult. Harald zur Hausen

Dedicated to Prof. Dr. Gerasimos Tsivgoulis,

who taught me that

“*Tensions are like a storm, they don’t last long.*

If again, it rains on a daily basis, get used to living with an umbrella.”

&

to Prof. Dr. Ioannis Matsoukas,

who introduced me to my mentor

TABLE OF CONTENTS

1. Introduction

Breast cancer is the most common cancer to affect women worldwide and accounts for approximately 30% of new female cancer diagnoses (Siegel, Kimberly et al. 2017). The survival rates differ globally and have been reported to be almost double for the developed countries compared to developing ones (Coleman, Quaresma et al. 2008). This difference is mainly attributed to enduring improvement of screening mammography, helping in early detection of breast cancer. In breast cancer, cells lose their normal regulation and multiply uncontrolled invading adjacent tissues. Eventually, after leaving the initial location, these cells travel to different body parts (such as liver, brain, lung etc.), known as metastasis, which is substantially the cause of breast cancer mortality. There are several kinds of breast cancer, most frequently the infiltrating ductal carcinoma originating from the milk ducts of the breast. Other types of invasive breast cancer are the infiltrating lobular carcinomas, medullary carcinomas, mucinous carcinomas, tubular carcinomas and several more (Akram, Iqbal et al. 2017).

- **Epidemiology and risk factors**

A number of factors are associated with an increased risk of breast cancer, such as sex, age, affluence, family history, breast conditions and endogenous estrogens, as well as hormonal and lifestyle factors (National Breast and Ovarian Cancer Center 2009) (**Table 1.1**). More specifically, women have 100-fold higher risk to develop breast cancer compared to men (Thomas 1993) and the incidence increases with age . In addition, several genes have been identified that play a significant role in the risk of breast

cancer development, including BRCA1, BRCA2, TP53, PTEN and CHEK2 gene, with mutations in these genes increasing the risk of breast cancer (Antoniou, Pharoah et al. 2003, Ghoussaini, Pharoah et al. 2013). Furthermore, giving birth to children reduces the risk of breast cancer about 30% depending on the number of deliveries and on breastfeeding period compared to women without children (Marianne, W. et al. 1990, Kelsey, Gammon et al. 1993), probably due to changes in breast epithelial cells because of lactation. These cells seem to be less susceptible to DNA damage, preventing the development of cancer (Gadducci, Biglia et al. 2005).

Table 1.1 Risk factors for breast cancer (adapted from Akhter et al. 2014).

Factors	Risk level
Age	Increase of risk, especially after 50
Sex	100x fold for women than in men
Age at menarche and menopause	Increase of risk for menarche before 14 and menopause after 54
Age at first full pregnancy	Double risk if first child after 30 3-fold increase without pregnancy
BRCA1 / BRCA2 gene mutation	5x fold increase of risk
High mammographic breast density	Increased risk
Diet and nutrition	Increased risk after high intake of saturated fat
Alcohol consumption	Excessive intake
Exposure to ionizing radiation	Abnormal exposure in females after age of 10

Interestingly, variations in nutrition and diet have attracted attention as possible evidence for the differences observed in breast cancer risk internationally. It has been suggested that the consumption of red meat (including beef, pork, lamb, goat, horse) increases the risk for carcinogenicity

(Bouvard, Loomis et al. 2015). Not only cancer, but also cardiovascular diseases and type 2 diabetes have been reported to be associated with long-term consumption of processed red meat (Alisson-Silva, Kawanishi et al. 2016). The role of red meat consumption in carcinogenesis has been attributed to compounds produced during cooking, such as heterocyclic amines and polycyclic aromatic hydrocarbons and others. However, the same compounds are produced when cooking fish and poultry, the consumption of which is not associated with an increased risk of cancer (Alisson-Silva, Kawanishi et al. 2016).

In addition to these studies, a new theory introducing a putative role of species-specific factors contributing to the etiology of certain diseases suggests that the consumption of red meat and dairy products originating from the species *Bos taurus* may explain a correlation with the global patterns of colon and breast cancer (zur Hausen and de Villiers 2015). According to the authors, in almost all high risk regions such us North America, most of the European countries and Australia, dairy cattle products are most consumed, while in low risk countries, such as Mongolia, Bolivia and India, we find different cattle species, such as Yak, Zebu and Zebu-Yak crossbreeds. This study also highlights the similarity in global incidence patters between colon and breast cancer (**Figure 1.1**). Yet, interesting differences are observed. More specifically, in India, breast cancer is increasing with higher rates compared to colon cancer and given that the Indian population does not consume beef meat, this increase could only be explained by the regular consumption of dairy products in recent decades, especially early in life. Furthermore, interesting differences are observed in Japan and Korea, where the milk consumption has been reported to increase after the Second World War and large quantities of beef meat were imported from the USA (zur Hausen and de Villiers 2015). Based on these observations, the authors suggested that dairy cattle may contain and transfer to humans a factor that is potentially involved in the development of breast and colon cancer, possibly as a result of various synergistic interactions (zur Hausen and de Villiers 2015, zur Hausen, Bund et al. 2017). The same theory has been suggested to

apply to neurodegenerative diseases, such as Multiple Sclerosis and Parkinson's disease, as well as to cardiovascular and autoimmune diseases.

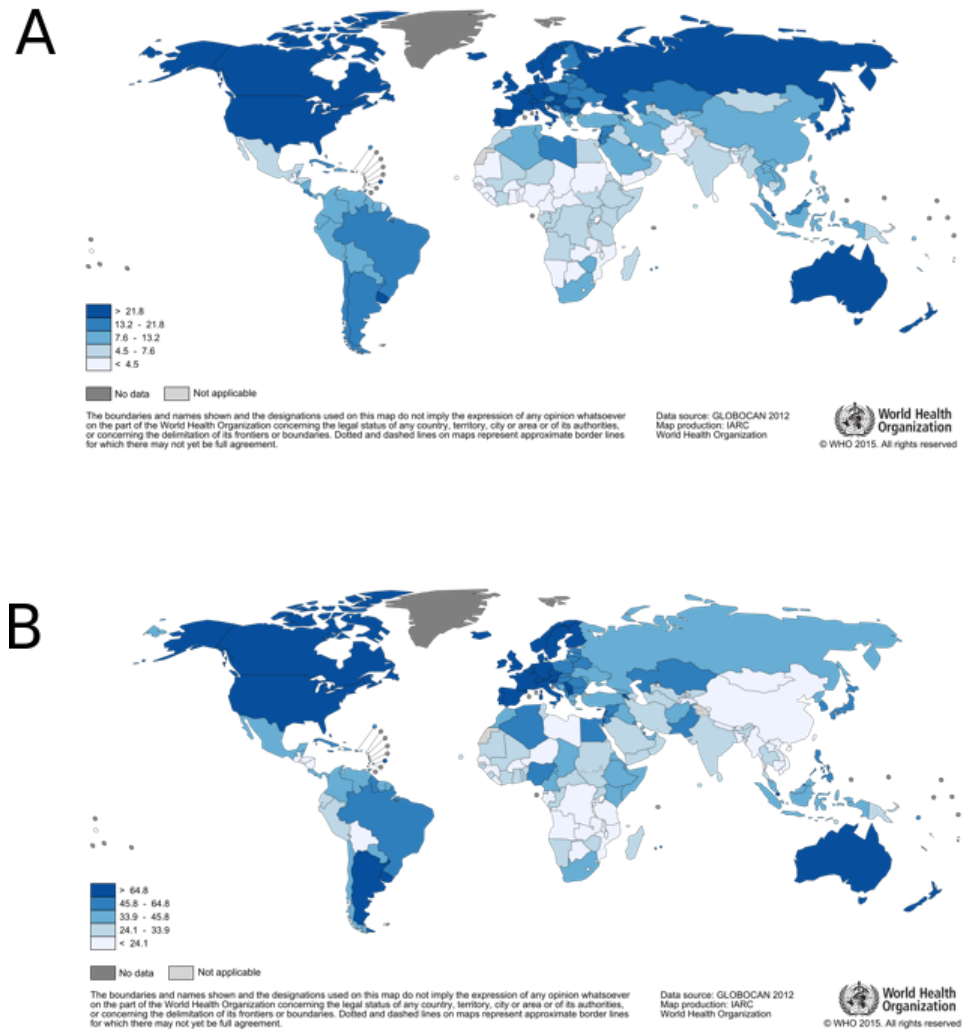


Figure 1.1 Global estimated incidence of colon (A) and breast cancer (B), GLOBOCAN 2012 (IARC). High risk regions for cancer incidence are shown in dark blue color and show similarity for the two types of cancer and the same applies for low risk counties indicated with light blue color.

A population study in Sweden revealed significantly lower risk of breast cancer for individuals with lactose-intolerance compared to their lactose-tolerant-parents and siblings. This was explained by the low consumption of

lactose-containing products of the lactose intolerant persons (Ji, Sundquist et al. 2015).

The theory for a potential tumor virus present in beef products which is not destroyed by overcooking was confirmed to an extend by the recent isolation of three “hamburger “ polyomaviruses from beef samples purchased in supermarkets of San Francisco, USA (Peretti, FitzGerald et al. 2015). Following a well-established approach of virion enrichment, rolling circle amplification and next generation sequencing, they successfully detected three polyomavirus species, as well as an adenovirus and a herpesvirus, implicating that the presence of such viruses in meat would potentially affect human health.

- **Isolation of bovine meat and milk factors (BMMFs)**

The study of epidemiological data and speculations that were raised based on a potential role of virus-like agents in the development of cancer and other diseases led the group of Prof. de Villiers and Prof. Dr. zur Hausen into the analysis of bovine serum samples, commercially available milk, as well as human blood and tissue samples from Multiple Sclerosis patients. This study resulted in the isolation of 18 circular single-stranded DNA agents (Funk, Gunst et al. 2014, Gunst, Zur Hausen et al. 2014, Lamberto, Gunst et al. 2014, Whitley, Gunst et al. 2014), which have been divided in four groups according to their nucleotide similarity to known DNA molecules (**Figure 1.2**). The isolates of Group 1 and Group 2 show similarity to Sphinx1.76 (acronym for **S**low **P**rogressive **H**idden **IN**fections of variable (**X**) latency) and Sphinx2.36 at the nucleotide level respectively. These represent episomes that have been isolated from brain and cultured neuronal cell lines of Creutzfeldt-Jakob disease (CKD) and sheep scrapie and show significant homology to *Acinetobacter* plasmids (Manuelidis 2011). Recently, the expression of the small Sphinx has been reported in neural cells and brain (Yeh, Gunasekharan et al. 2017). Due to the bovine origin of our Group 1 and

2 isolates, they have been called bovine meat and milk factors (BMMFs). Group 3 consists of isolates related to *Gemycircularviruses*, a group of single-stranded DNA viruses infecting plants and mammalians (Male, Kraberger et al. 2016). The isolate of Group 4 is distantly related to a plasmid of *Psychrobacter spec.*, which is found in a variety of sources (poultry, fish, sea water, clinical samples) and also as contaminants on complex media (reviewed in zur Hausen, Bund, and de Villiers 2017).

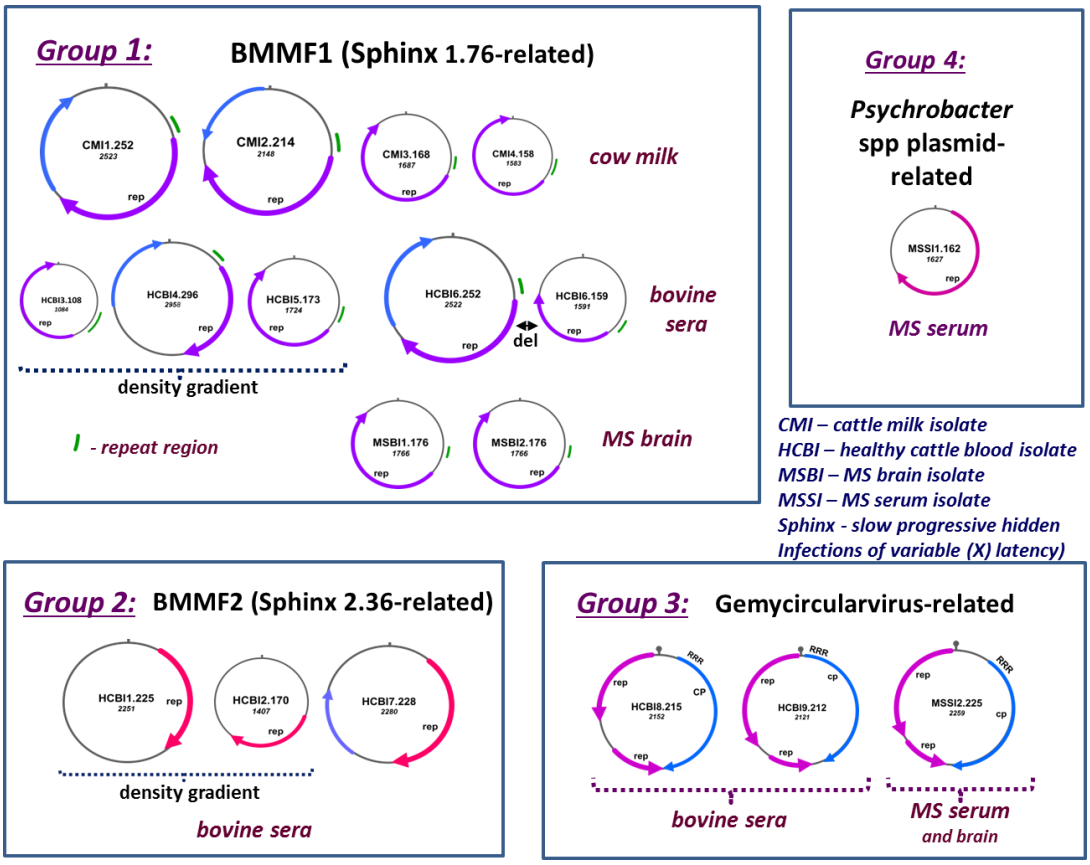


Figure 1.2 Schematic view of the novel episomal circular DNAs isolated from milk, bovine serum and tissue and blood samples from MS patients (adapted from zur Hausen, Bund and de Villiers 2017).

All isolates have an open reading frame (ORF) which encodes for a putative replication (Rep) protein and shares a high amino-acid similarity among the isolates (45-95%). The larger genomes include several additional

ORFs, the function of which is still under investigation. Furthermore, all isolates have an adenine and thymine (AT)-rich region and an iteron-like repeat region upstream of the rep protein, which is highly conserved, yet having single-nucleotide differences among them. These features are characteristic regions responsible for DNA replication initiation found in bacterial plasmids and eukaryotic replicons (Rajewska, Wegrzyn et al. 2012). The AT-rich region is located at the opening of the double helix where DNA synthesis starts. The iterons are responsible for binding the plasmid DNA replication initiator. The role of these complexes is important for the regulation and molecular mechanisms that are involved in replication.

The MSBI1.176 Rep protein shows significant homology to replication initiator proteins of the *Acinetobacter baumanii* plasmids (Whitley, Gunst et al. 2014). This bacterium is often related to nosocomial infections (Towner 2009). The presence of circular DNA in human brain, which is similar to an *Acinetobacter* plasmid, is an important observation that needs further investigation. Similar studies have already shown similarity of *Acinetobacter* to Sphinx2.36 (Longkumer, Kamireddy et al. 2013).

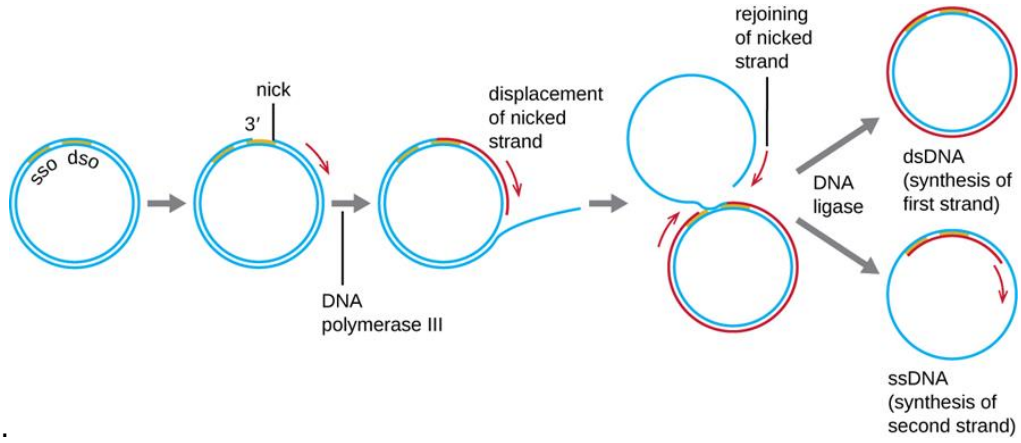
In order to understand a potential role of BMMFs in the pathogenesis of certain diseases, it was important to investigate BMMF replication in human cells and their ability to be transcribed and translated. General transcription of all isolates tested this far has been demonstrated in HEK293TT cells, as well as a helper effect in the replication of MSBI1.176 after co-transfection with CMI1.252 (Eilebrecht, Hotz-Wagenblatt et al. 2018). The latest observations raised questions whether complementation of individual, yet similar agents, is possibly responsible for their pathogenic role in human. In addition, this study established the presence of antibodies against the MSBI1.176 Rep protein in plasma of healthy donors suggesting that there has been an exposure towards BMMF earlier in life. Furthermore, it was demonstrated that MSBI1.176 and CMI1.252 isolates differentially regulate several genes, which have been suggested to be involved in carcinogenesis or neurodegenerative diseases.

- **Single-stranded DNA viruses**

Viral metagenomics have uncovered many novel types of single-stranded DNAs (ssDNAs) and have expanded the environmental distribution of known viral groups (Rosario, Duffy et al. 2012, Simmonds, Adams et al. 2017). Many ssDNA viruses identified through metagenomics encode a conserved replication-associated protein, so called circular Rep-encoding ssDNA viruses (CRESS-DNA viruses) (Rosario, Duffy et al. 2012). They belong to the families of plant pathogens *Geminiviridae* and *Nanoviridae,* but also to vertebrate pathogens like *Anelloviridae, Parvoviridae* and *Circoviridae* (Rosario, Duffy et al. 2012). During the last decade these viruses have been reported to be found in several environments including wastewater, sea waters, fecal samples, insects (Rosario, Schenck et al. 2015). The rapid evolution and high nucleotide substitution rates are characteristics of these viruses and have emerged as new taxonomic classification for this viral group (**Figure 1.3**).

The vast majority of CRESS DNA viruses replicates through Rolling Circle Replication (RCR) (Rosario, Duffy et al. 2012). Briefly, the initiation in RCR requires nicking of the double-strand origin (dso) by Rep and generation of a 3'-OH end, where the host DNA polymerases start the leading strand replication. Subsequently, elongation includes entension of the 3' terminus using the complementary strand as template. Once the viral strand is synthesized, Rep cleaves again at the reconstituted nick site and results in the formation of a new viral genome (**Figure 1.4**) (Rizvi, Choudhury et al. 2015, Ruiz-Maso, Macho et al. 2015)

Figure 1.3 Genetic diversity of CRESS-DNA viruses. Comparison of the Rep sequences of more than 600 CRESS-DNAs with 10 representative Rep sequences from viruses classified in the families *Gemonoviridae, Geminiviridae*, *Nanoviridae* and *Circiviridae,* and a group of alha satellites that are associated with geminiviruses or nanoviruses and representation with phylogenetic tree showing the similarity among the different viruses in amino acid level (Simmonds, Adams et al. 2017).



.

(https://teaching.ncl.ac.uk/bms/wiki/images/thumb/7/78/RollingCircleReplication.jpg/800px-RollingCircleReplication.jp)

Figure 1.4 Overview of rolling circle amplification. It starts with nicking of the double-strand origin (dso) by Rep and generation of a 3'-OH end, where the host DNA polymerases start the leading strand replication. Subsequently, elongation includes entension of the 3' terminus using the complementary strand as template. Once the viral strand is synthesized, Rep cleaves again at the reconstituted nick site which results subsequently in the formation of a new viral genome.

ssDNAs are often accompanied by subviral ssDNA molecules. Begomoviruses, for example, are associated with two major groups of satellite DNAs, which encapsidate and replicate with the help of the main virus, the alpha- and beta- satellites. The alpha-satellites can replicate autonomously, but they do not have a role in viral infection. Beta-satellites influence the pathogenicity of the virus by encoding a gene responsible for symptom development. In addition, there are reports of molecules smaller than 1 kb with unknown role, which are found in several types of viruses (Rosario, Duffy et al. 2012). Several reports document recombination events taking place between helper viruses and beta-satellites as well between alpha- and beta-satellites (Huang, Xie et al. 2013), while similar events have been documented for herpes simplex virus (Vlazny and Frenkel 1981) and varicella-zoster virus (Vlazny and Hyman 1985). Notably, the recombinant molecules of

approximately 700 nucleotides are capable of autonomous replication and interfere with the replication of the helper virus by defective interference.

We speculate that the isolated agents (**Figure 1.1**) belong the family of CRESS-DNA viruses, since they share important characteristics such as the replication protein and the AT-rich region followed by four direct repeats up stream of the putative Rep ORF.

- **Viral infections & potential role of glycoproteins as receptors**

Readdressing the hypothesis that the consumption of red meat might act as potential risk factor for the development of cancer, a compelling theory has been recently suggested (Alisson-Silva, Kawanishi et al. 2016). According to these authors a new mechanistic model could explain the tendency for risk related to the consumption of red meat; they propose that the incorporation of the sialic acid N-glycolylneuraminic acid (Neu5Gc) into human membranes leads to an immune response. Sialic acids in general have an important role in cell-cell interactions and are the link between the cell and the extracellular environment. Neu5Gc cannot be synthesized in humans due to an exon deletion in the enzyme cytidine monophospho-N-acetylneuraminic acid hydroxylase (CMAH), which is responsible for the production of Neu5Gc by its ancestor sialic acid N-acetylneuraminic acid (Neu5Ac). Thus, Neu5Gc enters the human tissues only externally through the consumption of food rich in Neu5Gc, especially red meat and milk products. It has been reported that anti-Neu5Gc antibodies appear during the first year of life (Taylor, Gregg et al. 2010). Neu5Gc integrated into cellular glycoproteins and gangliosides might be targeted by circulating anti-Neu5Gc antibodies, causing chronic human inflammation (“xenosialitis”), which promotes carcinogenesis and supports cancer progression (Samraj, Laubli et al. 2014). The presence of anti-Neu5Gc antibodies found in sera from cancer patients suggest that they could serve as tumor biomarkers and ELISA tests sialoglycan microarrays could be used as

screening tools for assessment of persons at risk and for monitoring therapeutic response.

A recent review summarizes why Neu5Gc glycoproteins might act as receptors for infection of BMMFs. Mainly, the authors report that it has already been demonstrated that several viruses bind specifically to Neu5Gc receptors, such as canine and fenine parvoviruses, polyomavirus type 9 and influenza and parainfluenza viruses. Furthermore, the authors highlight a potential relationship between anti-BMMF reactivity and Paul-Bunnell antibodies (Golaszewska, Kurowska et al. 2003), which recognize the Neu5Gc as immunogenic epitope in infectious mononucleosis, which has been reported to increase risk for Multiple Sclerosis (Haahr, Koch-Henriksen et al. 1995, Thacker, Mirzaei et al. 2006). In addition, they suggest that the conversion of Neu5Ac to Neu5Gc could potentially take place due to a Herpesvirus-specific O-glycosylating hydrolase and anti-BMMF reactivity could be increased after co-infection of cells with latently persistent BMMFs and EBV. Finally, anti-Neu5Gc antibodies appear during the weaning period, which is also the time of potential exposure of humans during the weaning period to BMMFs through the consumption of commercially available milk and dairy products.

- **Viruses and breast cancer**

The role of viral infections in cancer development has been debated over decades. Nowadays, it is established that a number of viruses are responsible for carcinogenesis: human papillomaviruses (HPVs) 16 and 18, Epstein-Barr virus (EBV), Kaposi's sarcoma-associated herpesvirus (KSHV), hepatitis B virus (HBV), hepatitis C virus (HCV), Merkel Cell polyomavirus (MCV) and human adult T-cell leukemia virus type 1 (HTLV-1) (White, Pagano et al. 2014). Research on the mechanisms of oncogenesis of these viruses is still ongoing and so far important knowledge has been gained. In general, oncogenesis can be generated a) by contribution of viral genes (e.g. HPV), b) by reduction of host immunity (e.g. human immunodeficiency virus, HIV) and

c) via chronic inflammation (e.g. HBV, HCV) (zur Hausen and de Villiers 2014, Gannon, Antonsson et al. 2018). However, the possibility of existence of yet unidentified agents or known ones for which no sufficient data exist at the moment and which might be associated with human cancer is under investigation.

The identification of mouse mammary tumor virus (MMTV) in late 1930s fueled the discussion of a potential role of viruses in breast cancer. The key candidate viruses for breast cancer and eventually the most studied are HPV, EBV, MMTV and bovine leukemia virus (BLV). A number of studies supports an association between viral infections and breast cancer (Lawson and Heng 2010, Glenn, Heng et al. 2012, Akhter, Ali Aziz et al. 2014, Lawson, Glenn et al. 2016, Nartey, Mazzanti et al. 2017, Lawson, Salmons et al. 2018). However, when these data were reviewed by other scientists, they failed to support an oncogenic viral infection being involved in the causality of the disease and do not justify a conclusion that any of the viruses mentioned above have a valuable role in the etiology of breast cancer. (Joshi and Buehring 2012, De Paoli and Carbone 2013, Akhter, Ali Aziz et al. 2014, Gannon, Antonsson et al. 2018). This debate might be a result of the relatively low abundance of viral DNA in human tissue and specifically within the final tumor, as well as the sample storage and preparation, potential laboratory contamination and the detection method used in each study.

- **Viral metagenomics**

Viral infections play a significant role in global health and the identification of novel pathogens is of great importance. During the last two decades, several outbreaks due to previously unknown viruses were reported, such the severe acute respiratory syndrome (SARS) coronavirus, the influenza H1N1 virus and the Ebola virus (Chiu 2013). Traditional methods of viral discovery included propagation of the virus in cell culture, serology or polymerase chain reaction (PCR) using consensus primers and hybridization

methods. However, these approaches require *a priori* knowledge of the identity of the infectious agent and have little or no value in identifying novel viruses. This led the researchers to “metagenomics”, an approach that does not require any prior knowledge about the investigating organisms (Handelsman 2004). The sequence-independent techniques includes random amplification techniques, such as the rolling circle amplification (RCA), which has been used in the detection of circular DNA viruses such as polyomaviruses, papillomaviruses and circoviruses (Niel, Diniz-Mendes et al. 2005, Schowalter, Pastrana et al. 2010, van der Meijden, Janssens et al. 2010, Dela Cruz, Giannitti et al. 2013), as well as the sequence-independent single-primer amplification (SISPA), which was used in the identification of Hepatitis E virus, Norwalk virus, Human astrovirus and Parvoviruses 2 and 3 (Mokili, Rohwer et al. 2012).

Next Generation Sequencing can be used for complete genome resequencing for mutation discovery in human genomes, transcriptome and small RNA sequencing for gene expression, alternative splicing and microRNA profiling, chromatin immunoprecipitation sequencing (ChIP-Seq) for mapping of protein-DNA interactions and sequencing of bisulfite-treated DNA for cytosine methylation in genomic DNA (Shendure and Ji 2008).

Three are the main steps involved in metagenomic analysis: a) sample preparation, b) high-throughput sequencing and c) bioinformatic analysis (Mokili, Rohwer et al. 2012).

a) *Sample preparation*

Detection sensitivity for the identification of novel pathogens depends on the reduction of the host background and the physical enrichment of the virus particles. Purification of virus capsid, filtration, ultracentrifugation and digestion with DNase digestion, as well as enrichment methods such as the RCA and the SISPA are often used for enrichment and degradation of host genomes (Mokili, Rohwer et al. 2012, Chiu 2013).

b) *High-throughput sequencing*

The selection of parameters in the sequencing platform is of critical importance in the detection of novel viruses. The generation of long sequence reads permits the generation of long contigs in *de novo* assembly, increasing the possibility of identification of sequences that correspond to unknown viruses. Furthermore, there must be sufficient generated reads for adequate sequencing depth (Tang and Chiu 2010, Chiu 2013).

c) *Bioinformatic analysis*

The key steps in analysis of the data obtained from NGS include the quality assessment and the removal of low-quality sequences (low quality corresponds to accuracy of base calling, it is filtered by Q score and a Q score of 30 refers to a base with a 1 in 1000 probability of being incorrect (Cliften 2015)) and host DNA, followed by either mapping to a known genome, which acts as reference, or by *de novo* (reference-free) assembly, which allows non-overlapping regions to be combined into one consensus sequence (Mokili, Rohwer et al. 2012). The generated contigs are then compared to the GenBank using Basic Local Alignment Search Tool (BLAST) (Datta, Budhauliya et al. 2015).

Aim of the thesis

Small circular DNA genomes have been isolated from cow milk, bovine serum and blood and tissue samples from Multiple Sclerosis patients (Funk, Gunst et al. 2014, Gunst, Zur Hausen et al. 2014, Lamberto, Gunst et al. 2014, Whitley, Gunst et al. 2014, Falida, Eilebrecht et al. 2017). Additionally, epidemiological data point at a relation between the consumption of milk and red meat and an increased incidence of certain types of cancer, such as colon, breast and prostate cancer as well neurodegenerative diseases like Parkinson's disease and Multiple Sclerosis. However, little is known about the potential role of the above-mentioned isolates during the development of such diseases.

The main aims of this study were:

1. to expand the work previously done regarding the isolation of episomal DNA agents from cow milk to different dairy products (e.g. yogurt, crème fraiche, card cheese, butter).

2. to identify novel episomal DNA agents as well as agents related to the already identified molecules in breast cancer tissue samples, using Next Generation Sequencing (NGS) of DNA.

3. to search for a permissive cell system for the identified bovine milk and meat factors (BMMFs). Epidemiological data suggest a potential link between the consumption of red meat and animal exposure and the incidence of Hodgkin's disease (Khuder, Mutgi et al. 1999, Epstein, Chang et al. 2015), while in the past 24 full-length Torque Teno viruses (a single stranded DNA virus) have been isolated from a single patient with Hodgkin's lymphoma (Jelcic, Hotz-Wagenblatt et al. 2004). Therefore, two Hodgkin's lymphoma cell lines, the L-428 and L-1236, were selected for transfection with the MS brain isolate MSBI1.176 and they were analyzed regarding the replication of the genome. In

addition, co-transfection experiments with two different genomes were performed in order to test for a helper or interfering effect over replication, since it has already been reported that while replication of MSBI1.176 was abortive in HEK293TT cells, co-transfection with CMI1.252 (cow milk isolate) resulted in almost stable MSBI1.176 replication until 14 days post transfection. In addition, it has been reported that incorporation of N-Glycolylneuraminic acid (Neu5Gc) into glycoproteins of human cells leads to chronic inflammation which potentially promotes cancer and other diseases (Alisson-Silva, Kawanishi et al. 2016). In order to test an involvement of Neu5Gc in the permissiveness of human cells for BMMFs and establish an infection system for BMMFs, infection assays were performed using MSBI1.176 positive supernatants of MSBI1.176-transfected cells with and without addition of external Neu5Gc.

4. to analyze a potential link of BMMFs to breast cancer or to Hodgkin's disease etiology and to assess a potential immunological response to those agents, specifically in patients suffering from the respective disease. In the latter case, elevated antibody titers directed against BMMF gene products in sera of disease patients would be expected compared to healthy controls. Therefore, sera from breast cancer and Hodgkin's disease patients were analyzed by ELISA assays in order to identify a potential immune reaction of those patients against the Rep protein of MSBI1.176.

2. Materials and Methods

2.1 Chemicals

All chemicals were purchased from Merck (Darmstadt), Thermo-Fischer Scientific (Germany), Sigma-Aldrich (USA), Carl Roth (Karlsruhe), Gerbu Biotechnik (Heidelberg), Fisher Chemical (Loughborough) and QIAGEN Sciences (USA) in molecular biology grade, except where otherwise specified. Disposables were purchased from Greiner Bio-One (Germany), nerbe plus (Germany) and Eppendorf (Germany).

2.2 Enzymes, reagents and kits

Table 2.1 List of enzymes, reagents and kits.

PRODUCT	**VENDOR**
96-well MaxiSorp™ ELISA plates	Thermo Scientific
Agilent DNA 12000 Kit	Agilent Technologies
AmpliTaq Gold DNA Polymerase with Gold Buffer and $MgCl_2$	Applied Biosystems, Thermo Fisher Scientific
CellTiter-Glo® Luminescent Cell Viability Assay	Promega
Clarity™ Western ECL Substrate	Bio-Rad
Click-iT™ Edu Plus Alexa Fluor™ 647 Imaging Kit	Thermo Scientific
dNTP Mixture	Takara Bio Inc.
Exo-Resistant Random Primer	Thermo Scientific
FastAP Thermosensitive Alkaline Phosphatase (1 U / µl)	Thermo Scientific
GeneRuler™ DNA Ladder Mix, ready-to-use	Thermo Scientific
Glycogen	Invitrogen, by life Technologies
NEB 5-alpha Competent *E. coli* cells	New England BioLabs
N-glycolylneuraminic acid (Neu5Gc)	Sigma Aldich
Nucleofector® Kit V	LONZA

PfuUltra II Hotstart PCR Master Mix, 100 rxn	Agilent Technologies
phi29 DNA Polymerase	New England BioLabs
phi29 DNA Polymerase Reaction Buffer Pack	New England BioLabs
Plasmid-Safe™ ATP-Dependent DNase	Epicentre® Technologies
Proteinase K	
Restore™ PLUS Western Blot Stripping Buffer	Thermo Scientific
Restriction enzymes	Thermo Scientific
Restriction enzymes	New England BioLabs
Skim milk powder	GERBU Biotechnik
Superblock® T20 (PBS) Buffer	Thermo Scientific
SYBR Green PCR Master Mix	Applied Biosystems, Thermo Fisher Scientific
T4 DNA Ligase, 5 Weiss U/µl	Thermo Scientific
T4 Polynucleotide Kinase with Buffer	New England BioLabs
TA Cloning® Kit with One Shot® TOP10F´ Chemically Competent *E.coli*	Invitrogen
TaKaRa LA Taq® with GC Buffer	Takara Bio Inc.
TMB ELISA Substrate Kit	Thermo Scientific
Trans-Blot® Turbo™ Transfer System	Bio-Rad
TruPAGE SDS Running Buffer, 20x	Sigma Aldrich
TruPAGE™ Precast Gels 4-20%	Sigma Aldrich
Trypan Blue Solution, 0.4%	Thermo Scientific
Venor®GeM Classic	Minerva Biolabs

Enzymes were used according to the manufacturer's protocol. DNA extractions from tissue samples and cultured cells were performed with DNeasy Blood & Tissue Kit (QIAGEN), QIAamp® DNA Blood Mini Kit (QIAGEN), QIAprep® Spin Miniprep Kit (QIAGEN) and Plasmid Mini AX kit (A&A Biotechnology). Cell lysate homogenization for RNA isolation was done with the QIAshredder homogenizer (QIAGEN). For DNA extraction from agarose gels and PCR clean-up the NucleoSpin® Gel and PCR Clean-Up kit (MACHEREY-NAGEL) was used. Plasmid DNA purification from transformed

bacteria was done with the NucleoBond® Xtra Midi kit (MACHEREY-NAGEL) and the *AccuPrep*® Plasmid Mini Extraction Kit (BIONEER Corp.).

2.3 Synthetic oligonucleotides

All primers for polymerase chain reactions and sequencing were purchased from Sigma-Aldrich.

Table 2.2 List of primers used for PCR and qPCR.

PCR		
Primer name	**Forward (5´→ 3´)**	**Reverse (5´→ 3´)**
7717N_cont_80_16	AATGTGGCAAAAGGCGTTAC	GTGACCCATTCGTACTGCTTG
AL_CQ 3p AL_CQ 5p	CTTCTAGGTTATAACTTGCATTA	TCACAGAACAACGACTAATTC
Bc_con 3p Bc_con 5p	ATTGATTAAAGCATTATCTTT	GCCAGTTATAACCTAGA
Dx 3p Dx 5p	ATTCCAAGTTATAACTGGCATTG	TAACAGAACAGCGTTTGATTATG
FU_BF 3p FU_BF 5p	GGTCTAGGTTGTAGCTCG	TAGTGGAACAACGTTTGATTC
LF271 LR2249	ATTGCCCCTCGACTAAGCAAGC	TGTACGAAAGCTAGCACCACCG
M1sg613	AGATGCGAGTTATTGAATTAG	CTAAGCAAGCTTAGTCGAG
mtDNA sense mtDNA antisense	ACCACCTCTTGCTCAGCCTA	CATGGGCTACACCTTGACCT
mt2DNA sense mt2DNA	ACACTCATCGCCCTTACCAC	GTTGATGCAGAGTGGGGTTT

antisense		
mt3DNA sense mt3DNA antisense	GCTAAACCTAGCCCCAAACC	TTGGCTCTCCTTGCAAAGTT
mt4DNA sense mt4DNA antisense	TATCCGCCATCCCATACATT	GGTGATTCCTAGGGGGTTGT
Oa Ob	CTAGACGAGTTTAGAAAGCG	CTCAATAATTGGTGTTTTGCC
Oc Od	ACCAATTATTGAGCTAGACGAG	GTTTTGCCTGTGCTACGCC
Oe Of	GCACGAGTGGGTTACATCG	ACCCAACTGATCTTCAGCATCT
pNLF1-N sense pNLF1-N antisense	CAGGGAGGTGTGTCCAGTTT	GCCATAGTGCAGGATCACCT
pUC19	ACCGGAGCTGAATGAAGCCATACC	TCCCAACGATCAAGGCGAGTTACA
pUC19 sense pUC19 antisense	AGTTGGCCGCAGTGTTATCA	TCCTTGAGAGTTTTCGCCCC
sg613.L4.7/F,R sg613.L4.9/R	TAGATACTGAATACCCCTACG	CGTAGGGGTATTCAGTATC CGTAGGGGTATTCAGTATCTA
Nn (Sphinx 1.76) Xn (Sphinx 1.76)	GGATTAATGCCAATGATCC	CTTTGCCTGTTTCTCTCG
No (Sphinx 1.76) Xo (Sphinx 1.76)	GAGGACGAATTAATATTACAAGTC	GTTCTCGCTTTTCTTGGTAA
qPCR		
beta actin gene sense	CCACCATGTACCCTGGCATT	ATCTGAGGAGGGAAGGGGAC

beta actin gene antisense		
CMI1.252-specific sense CMI1.252-specific antisense	ACAAGCTAGAAGAATTTGGCGT G	TGGCACAGTAGAGCTTTCATCA
MSBI1.176 lin/circ sense MSBI1.176 lin/circ antisense	ACCCCTACGTTTACCTTGCG	CCTTTGCCTGTTTCTCTCGC
MSBI1.176-DpnI sense MSBI1.176-DpnI antisense	CCCACACAGCAAGGCATACA	CCTTTGCCTGTTTCTCTCGC
MSBI1.176-specific sense MSBI1.176-specific antisense	ACGGGTAGGCTTGCTTATTTGA	TGTATGCCTTGCTGTGTGGG
pUC19 sense pUC19 antisense	AGTTGGCCGCAGTGTTATC	TCCTTGAGAGTTTTCGCCCC
pUC19-specific sense pUC19-specific antisense	GCTCAGTGGAACGAAAACTCAC	AGCCCTCCCGTATCGTAGTT

2.4 Plasmids & DNA genomes

Table 2.3 Vectors used and their application.

Vector	**Application**	**Vendor**
pCR2.1	Standard cloning of PCR products	Invitrogen, by life Technologies
pUC19	Sub-cloning of restriction-digested DNA fragments	New England Biolabs
pcDNA™3.1 (-)	Cloning of proteins for over-expression	Invitrogen

pNLF1-N	Control in protocol establishment for isolation of plasmid DNA from human tissue	Promega
pZsGreen	Transfection efficiency control	Clontech
pmaxGFP®	Transfection efficiency control	LONZA

Table 2.4 DNA genomes from the lab mentioned in the Results part.

Genome	**Full name**	**Accession Number (GenBank)**
CMI1.252	Cow Milk Isolate 1	LK931487
CMI2.214	Cow Milk Isolate 2	LK931488
CMI3.168	Cow Milk Isolate 3	LK931489
CMI4.158	Cow Milk Isolate 4	LK931490
CMI5.170 / CMI5.240	Cow Milk Isolate 5	LT715554, LT715555
HCBI3.108 / HCBI4.296 HCBI5.173 / HCB1.6.159 HCBI6.252	Healthy Cow Blood Isolate 3,4,5,6	LK931495, LK931496 LK931497, LK931494 LK931493
MSBI1.176	Multiple Sclerosis Brain Isolate 1	LK931491
MSBI2.176	Multiple Sclerosis Brain Isolate 2	LK931492

2.5 Standard media (E. coli)

LB (Luria Bertani): 0.5% (w/v) tryptone

1% (w/v) yeast extract

0.5% (w/v) NaCl in sterile water.

After autoclaving ampicillin (100 µg / ml) was added.

For preparation of media plates the LB medium was supplemented with 1% agar and sterilized. 0.2% X-Gal and 0.2% IPTG were added and after mixing, the plates were poured.

2.6 Standard solutions and buffers

Chloroform:Isoamyl alcohol 24:1 (CIA)	240 ml Chloroform
	10 ml Isoamyl alcohol
TE 1x:	10 mM Tris base
	1 mM EDTA
	pH 8
2x PK Buffer	0.2 M Tris-HCl, pH 7.5
	25 mM EDTA
	0.3 M NaCl
	2% SDS
Proteinase K dilution	20 mg / mL in sterile water
3 M sodium acetate solution pH 5.2	40.8 g sodium acetate
	pH 5.2
	In a final volume of 100 ml

Ammonium acetate 7.5 M:	57.81 g ammonium acetate / 100 ml
1 x EP Buffer:	40 mM Tris Base
	5 mM sodium acetate
	1 mM EDTA
	pH 7.8
DNA loading dye:	0.5% Orange G
	50% glycerol
Freezing medium:	17% RPMI-1640 Medium with L-glutamine and sodium bicarbonate
	20% FCS
	10% DMSO
5 x Lämmli Buffer	250 mM Tris-HCl pH 6.8
	8% SDS
	30% glycerol
	0.2% bromophenol blue
2.5 Lämmli buffer	4.5 ml 5 x Lämmli
	4.5 ml sterile H_2O
	1 ml 2-Mercaptoethanol

ELISA reagents

Coating buffer: PBS 8M Urea	8 M Urea 100 mM $NaH_2PO_4x2H_2O$ 10 mM Tris-HCl pH 8 In a final volume of 500 ml
Wash buffer: PBS/Tween	1 x PBS pH 7.2 0.1% (v/v) Tween 20
Stop solution	1 M sulfuric acid 8 M acetic acid

Western blot reagents

Ponceau Staining Solution	0.5% (w/v) Ponceau S 1% acetic acid
Blocking buffer	5% (w/v) skim milk powder 1xPBS
Washing buffer	0.1% Tween 20 1xTBS

TBS 1x	150 mM NaCl 20 mM Tris-HCl pH 7.6 In a total volume of 1 L

2.7 Softwares / web interfaces

HUSAR (Heidelberg Unix Sequence Analysis Resources)	Web site: https://www.dkfz.de/gpcf/hs_husar.html
NEBtools™ Tm Calculator	Web site: http://tmcalculator.neb.com/#!/main
Primer BLAST (NCBI)	Web site: https://www.ncbi.nlm.nih.gov/tools/primer-blast/
Quantity One 1-D Analysis 4.6.9	Bio-Rad
GIMP 2.8.22	GIMP
Image Lab™ Software 5.2.1	Bio-Rad
QuantStudio™ Design & Analysis Software v1.4	Applied biosystems, ThermoFisher Scientific
Clone Manager v6	Sci-Ed Software
SigmaPlot 13.0	Systat Software Inc.
PuTTY 0.70	PuTTY

2.8 Molecular Biology Methods

2.8.1 Phenol-isoamyl-chloroform extraction of DNA & ethanol precipitation

For the lysis step 12 ml of the sample, 12 ml 2 x PK Buffer and 500 µl Proteinase K dilution were added in a 50 ml falcon and the mixture was rotated overnight at 37 ^{o}C. Subsequently, the sample was mixed with 1 volume of phenol for 20 min and centrifuged at 5000 rpm for 15 min at room temperature. The aqueous phase was transferred to a fresh falcon tube and the previous step was repeated. The aqueous phase was transferred to a fresh falcon tube and mixed with ½-volume phenol and ½-volume CIA for 20 min, followed by centrifugation at 5000 rpm for 15 min at room temperature. The aqueous phase was transferred to a fresh falcon tube, mixed with 1 volume CIA for 20 min and centrifuged at 5000 rpm for 15 min at room temperature. The aqueous phase was transferred to a fresh tube and 2 volumes of absolute ethanol, $1/10$ of the volume of 3 M sodium acetate solution and 1 µl glycogen were added. For DNA precipitation the mixture was incubated for at least 2h at -20 ^{o}C. Consequently, the mixture was centrifuged for 45 min at 12000 rpm at 4 ^{o}C, the supernatant was discarded and the pellet was rinsed with 70% ethanol. Centrifugation was repeated once and the DNA pellet was air-dried and dissolved in an appropriate volume of TE buffer.

2.8.2 Purification of extracted DNA

For the purification of the isolated DNA, the NucleoSpin® Gel and PCR Clean-Up kit (MACHEREY-NAGEL) was used following the manufacturer's protocol for PCR clean-up. The elution step was repeated 3 times with 30 µl NE buffer heated at 70 ^{o}C.

2.8.3 Rolling circle amplification (RCA)

Rolling Circle Amplification by the phi29 DNA Polymerase was used to specifically amplify circular DNA molecules. For the RCA the phi29 DNA Polymerase Reaction (New England BioLabs) was used according to the following protocol: 50 ng total DNA were placed into a PCR tube containing 1x phi29 DNA Polymerase Reaction Buffer, 50 μM Exo-Resistant Random Primer (Thermo Scientific), 0,01 mg BSA and the total volume was adjusted with H_2O to 10 μl. The mixture was heated to 95 ^oC for 3 min and cooled to room temperature. The reaction was brought to a final volume of 20 μl containing 1x phi29 DNA Polymerase Reaction Buffer, 0.004 mg BSA, 0.4 mM dNTP Mixture (Takara Bio Inc.) and 10 U phi29Pol (New England BioLabs) and it was incubated at 30 ^oC for 18 h followed by 10 min incubation at 65 ^oC.

2.8.4 Polymerase chain reaction (PCR)

A standard PCR was perfomed using the TaKaRa LA Taq® in 1x GC Buffer (Takara Bio Inc.) in a final volume of 50 μl containing 100-200 ng DNA template, 10 μM of each primer (forward and reverse), 8 μl dNTP mix (2.5 mM each) and 2.5 U LA Taq® Polymerase. The reaction was run in a thermocycler according to the protocol bellow (Table 2.5).

Table 2.5 Standard PCR amplification program.

Time	Temperature	Step	
1 min	95 ^oC	denaturation	
30 sec	95 ^oC	denaturation	30-40 x
1 min	T_m - 5 ^oC	annealing	
1 - 4 min	72 ^oC	elongation	
10 min	72 ^oC	final extension	

For proof reading PCR, proof reading DNA polymerase (PfuUltra II Hotstart PCR Master Mix, 100 rxn, Agilent Technologies) was used in order to certify definite amplification.

2.8.5 Agarose Gel Electrophoresis

DNA was analyzed by agarose gel electrophoresis based on the molecular weight / electro mobility. 1/6 volume of loading dye was mixed with each sample and run on a 1 - 2.5% (w / v) agarose gel. The agarose percentage was adjusted according to the size of the DNA fragments to be separated. The gel was prepared in 1x EP buffer, which was also used as running buffer. The gel was run at 90 V. The size was estimated based on the GeneRuler™ DNA Ladder Mix, ready-to-use (Thermo Scientific) **(Fig. 2.1)**. The gel was stained in 1x EP buffer containing ethidium bromide for at least 15 min and stained DNA fragments were visualized by UV-light.

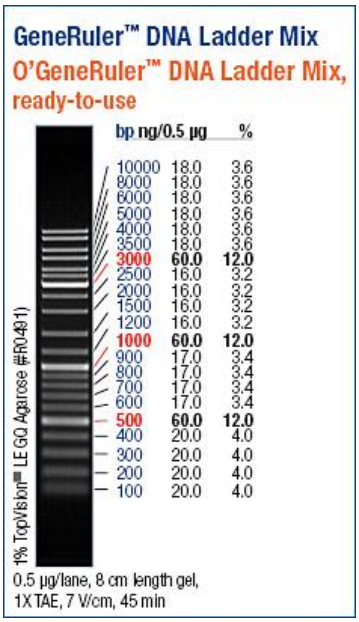


Figure 2.1 DNA ladder for agarose gel electrophoresis (GeneRuler™ DNA Ladder Mix, ready-to-use, Thermo Scientific).

2.8.6 DNA extraction from agarose gels

For subcloning or further processing, DNA fragments were extracted from agarose gels after electrophoresis. Therefore, the desired band was cut with a scalpel under UV-light. For the extraction of DNA, the NucleoSpin® Gel and PCR Clean-Up kit (MACHEREY-NAGEL) was used according to the manufacturer's protocol.

2.8.7 Ligation of DNA fragments

2.8.7.1 Ligation with the TA cloning kit (Invitrogen)

DNA fragments obtained from PCR using the Taq DNA polymerase contained a TA overhang which made them suitable for subcloning with the TA Cloning® Kit (Invitrogen). The amount of extracted DNA insert used for ligation with 50 ng of pCR2.1® vector was calculated based on the equation:

$$X\,ng\,insert = \frac{(Y\ bp\ insert)\left({}^{1}/_{3} * 50\ ng\ pCR2.1®\ vector\right)}{(size\ in\ bp\ of\ \text{pCR2.1® vector} \sim 3900)},$$

where X ng is the amount of the DNA isert to be ligated and Y is the number of base pairs of the insert for a ratio of 1:3 (vector:insert) (adapted from TA Cloning® kit, User manual).

For a standard ligation reaction in a total volume of 10 µl, the manufacturer's instructions were followed.

2.8.7.2 Ligation with other vectors

For the ligation of DNA fragments in vectors other than the pCR2.1, a 5´-dephosphorylation step of the linearized vector was preceded in order to avoid religation of the vector. Therefore, 2 U FastAP (Thermo Scientific) were added directly after restriction digestion with the corresponding enyzme and the mixture was incubated at 37°C for 15 min, followed by a 5 min incubation at 75°C. The insert DNA was restricted with the same restriction enzyme and purified from an agarose gel (see 2.7.6), before the ligation reaction was started.

The ligation was performed overnight at room temperature using the T4 DNA ligase (Thermo Scientific) in a final volume of 50 µl in 1x T4 DNA ligase buffer. The amounts of vector and DNA were calculated according to the equation shown in 2.7.7. The enzyme was inactivated by 15 min incubation at 65°C.

2.8.8 Transformation of competent E. coli cells

The One Shot® TOP10F´ chemically competent *E.coli* cells have been used for transformation of palsmids containing subcloned vectors, unless it is indicated differently. The transformation was performed according to manufacturer's instructions and the cells were plated on LB Medium plates containing IPTG, X-Gal and Ampicillin. They were incubated overnight at 37^{o}C.

2.8.9 Plasmid DNA isolation from E. coli cells

Starter bacteria cultures were inoculated by single colonies obtained after transformation and cultured overnight at 37 ^{o}C on a shaker at 130 rpm. The *AccuPrep*® Plasmid Mini Extraction Kit (BIONEER Corp.) was used for plasmid DNA purification according to manufacturer's protocol. The final elution volume was 50 µl. For midi scale plasmid preparations, the NucleoBond® Xtra Midi kit (MACHEREY-NAGEL) was used according to manufacturer's instruction. Therefore, cultures of 400 ml inoculated with a 1:1000 dilution of the corresponding starter cultures were incubated overnight at 37 ^{o}C on a shaker at 150 rpm. The isolated plasmid was dissolved in an appropriate volume of TE buffer. The amount (A_{260}) and purity (A_{260} / A_{280}) of the isolated DNA were measured on a UV Spectrophotometer.

2.8.10 DNA restriction enzyme digestion

For subcloning of DNA fragments and verification of generated plasmid DNAs, DNA was restricted using restriction endonucleases. Therefore, the desired amount of DNA was digested at 37^{o}C by the appropriate amount of restriction enzyme, which was determined by the equation:

$$\frac{U}{h} = \frac{L_R}{L_v} \cdot \frac{N_v}{N_R} \cdot m_v \ ,$$

where U = Units of enzyme

h = hour

L_R = size of reference genome (λ = 48.5 kb)

L_v = size of DNA of interest

N_v = number of restriction sites in DNA of interest

N_R = number of restriction sites in reference genome

m_v = mass of DNA of interest

2.8.11 Sequencing of plasmid DNA

The sequence of the isolated plasmids was determined by Sanger Sequencing from the GATC Biotech company. For a sequencing reaction, a total volume of 30 µl purified DNA with concentration of 50-70 ng / µl were used. The obtained sequences were analyzed using the HUSAR platform and the Clone Manager software (see 2.6).

2.8.12 Sub-cloning of pCR®2.1cloned genomes into pUC19

For transfection experiments, it was necessary to prepare circularized BMMF genomes lacking the residual vector fragments. The corresponding DNA had to transferred from pCR2.1 vector into the pUC19 vector bz using a unique restriction site present in both the BMMF genome and the multiple cloning site if the pUC19 vector. Therefore, the full pCR2.1-cloned BMMF genome was first amplified bz PCR using the same primers as for cloning into the pCR2.1 (NnXn primers, see Appendix). The PCR reaction was performed with 100 ng DNA template using the program shown in Table 2.6.

Table 2.6 PCR program for sub-cloning of pCR®2.1clone genome into pUC19.

Time	Temperature	Step	
2 min	95 °C	denaturation	
45 sec	95 °C	denaturation	x 35 cycles
45 sec	58 °C	annealing	
1,5 min	72 °C	elongation	
7 min	72 °C	final extension	

Immediately after PCR, the DNA template was digested with 25 U DpnI for 2 h at 37 °C. After agarose gel electrophoresis (see 2.7.5.), the band of interest was cut, extracted (see 2.7.6) in 45 µl sterile water followed by a phosphorylation step in 1x Polynucleotide Kinase Buffer (New England Biolabs) using 10 U T4 Polynucleotide Kinase (PNK, New England Biolabs) for 1 h at 37 °C. After inactivation of the PNK at 65 °C for 15 min, the DNA was purified by addition of 500 µl extraction equilibration buffer of the NucleoSpin® Gel and PCR Clean-Up kit (MACHEREY-NAGEL) and according to the manufacturer's instructions. Next, the resulting blunt-end DNA fragment was circularized by self-ligation. The self-ligation was performed with the full amount of purified DNA using 10 U T4 Ligase (Thermo Scientific) in 1x T4 ligase buffer in a final volume of 40 µl and overnight incubation at room temperature. Subsequently the sample was subjected to RCA (see 2.7.3) followed by overnight restriction digestion at 37 °C with the appropriate restriction enzyme in a final volume of 80 µl (see 2.7.10). 2 µg of pUC19 vector were also linearized with the same enzyme in a total volume of 20 µl. After agarose gel electrophoresis (see 2.7.5), the restricted RCA product was cut and extracted from the gel while 2 µl of the pUC19 linearized mixture were added during gel extraction. The DNA was extracted in 40 µl sterile water followed by overnight ligation with 10 U T4 Ligase (Thermo Scientific) at 37 °C. The enzyme was inactivated by a 15 min incubation at 65 °C. The resulting ligation product was used for transformation of chemically competent bacteria (see 2.7.8) followed by standard mini plasmid preparation (see 2.7.9) and sequencing (see 2.7.11).

2.8.13 Preparation of circularized genomes

For the generation of large amounts of circularized BMMF genomes for cell culture transfetcions, 250 µg pUC19-cloned BMMF genome were digested overnight at 37 ^{o}C with the appropriate restriction enzyme used for cloning in a total volume of 500 µl of the corresponding buffer (see 2.7.10). After agarose gel electrophoresis (see 2.7.5), the band of interest was cut, extracted and eluted from the gel (see 2.7.6). 50 µg of the linear BMMF DNA were ligated using 30 U T4 Ligase (Thermo Scientific) in 1x T4 DNA Ligase Buffer (Thermo Scientific) in a total volume of 9 ml at 16 ^{o}C for >50 h. 4.5 ml ammonium acetate (7.5 M) and 35 ml 95% ethanol were added before mixing for overnight precipitation of the genome at 4 ^{o}C. After 1h centrifugation at 5000 rpm and 4 ^{o}C, the DNA pellet was rinsed with 70% ethanol and dissolved in an appropriate volume of TE buffer. The efficacy of the self-ligation was tested by restriction digestion with two different restriction enzymes.

2.8.14 Quantitative PCR (qPCR)

qPCR analyses were performed on a QuantStudio^{TM} 5 System instrument type (Applied Biosystems) with default conditions according to manufacturer's insructions for 96-Well block type with standard curve and SYBR Green Reagents chemistry.10-fold dilution series of the template were prepared for running the standard curve. 200 ng of each sample were used for the reaction and each reaction was run in triplicate to measure reproducibility. 10 µM of each primer were mixed with 25 µl SYBR Green Master Mix (Applied Biosystems) and the total reaction volume was 50 µl. The data analysis was done with the QuantStudio^{TM} Design & Analysis Software.

2.9 Breast cancer tissue analysis by Next Generation Sequencing

2.9.1 DNA extraction from breast tissue samples

Four DNA extraction methods were compared for the isolation of circular DNA molecules using approximately 25 mg of placenta tissue, to which known

amounts of plasmid vector (pUC19, 2700 bp and Nanoluc, 5900 bp) were added before starting the isolation procedure.

a. Phenol-isoamyl-chloroform DNA extraction (isolation of total DNA)

   Approximately 25 mg placenta tissue were cut and placed into an Eppendorf tube containing 1.5 ml absolute ethanol. The sample was mixed by overnight rotation at room temperature. After a 5 min centrifugation at 12000 rpm, the ethanol was removed and a defined amount of vector DNA was added to the pellet, which was subsequently lyophilized. Following, a standard phenol-isoamyl-chloroform DNA extraction was performed, followed by an ethanol precipitation (see.2.7.1). The DNA pellet was dissolved in an appropriate volume of TE buffer.

b. DNeasy Blood & Tissue kit (QIAGEN) (isolation of total DNA)

c. QIAprep® Spin Miniprep Kit (QIAGEN) (isolation of plasmid DNA)

d. Plasmid Mini AX kit (A&A Biotechnology) (isolation of plasmid DNA)

All three kits were used according to manufacturer's protocol. For the plasmid kits, an additional step was added at the beginning of the DNA isolation. The mixture of placenta tissue and vector was rotated overnight at 37 ^{o}C with 250 µl 2x PK Buffer and 10 µl Proteinase K dilution in a total volume of 500 µl. The final elution was performed in 50 µl elution buffer or TE buffer.

2.9.2 Plasmid-Safe™ ATP-Dependent DNase digestion

In order to enrich DNA samples obtained from different types of tissue for circular DNA molecules, the samples were treated with Plasmid-Safe™ ATP-Dependent Dnase, an enzyme which selectively digests linear DNA molecules. Therefore, the DNA samples were subjected to RCA (see 2.7.3) followed by digestion with Plasmid-Safe™ ATP-Dependent DNase (Epicentre® Technologies). 2 µg DNA were digested according to manufacturer's protocol (Table 2.7) for 30 min at 37 ^{o}C. The enzyme was inactivated at 70 ^{o}C for 30 min and the DNA was subjected to ethanol precipitation (see 2.7.1) before being dissolved in an appropriate volume of TE buffer.

Table 2.7 Plasmid-Safe DNase digestion protocol.

Component	Amount
DNA	2 µg
25 mM Reaction Buffer	2 µl
Plasmid-Safe DNase (10 U)	1 µl
Sterile water	x µl
Total volume	50 µl

2.9.3 Quality control before library generation

The quality of the resulting DNA samples was assessed before library generation using the Agilent DNA 12000 Kit (Agilent Technologies) according to the manufacturer's protocol on a Agilent 2100 Bioanalyzer (Agilent Technologies).

2.9.4 Library preparation & sequencing

The libraries for NGS of DNA were prepared using the TruSeq Nano (Illumina) protocol for formalin-fixed paraffin-embedded (FFPE) samples according to the manufacturer's instructions. All libraries were barcoded in order to distinguish the different samples, pooled and sequenced in one lane. Both the generation of the libraries and the sequencing were performed by the Genomics Core Facility of the DKFZ.

2.9.5 Bioinformatical analysis

Table 2.8 Programs used for the bioinformatical analysis.

Program / custom pipeline	**Version**	**Application**	**Reference (Ref or PMID or URL)**
EvalRSeq following 5 tools were used:	1.2	Filtering and mapping	
Fastq_quality filter and artifacts filter from the FASTX Toolkit	0.0.13	Quality, artifacts filtering	http://hannonlab.cshl.edu/fastx_toolkit/index.html
Homertools	4.7	Removal of PolyA and reads < 17 bases	20513432
Cutadapt	1.71	Removal of Adapter sequences	http://dx.doi.org/10.14806/ej.17.1.200
CollectRnaSeqMetrics from Picard tools	1.78	Quality check of RNA-Seq data	https://broadinstitute.github.io/picard/
Tophat2	2.014	Mapping of reads to the human genome 38, splice aware	23618408
Other tools used:			
bedtools bam to fastq	2.17.0	Converts a BAM file into a FASTQ file	20112078
Bowtie2	2.1.0	Alignment of sequencing reads to long virus reference sequences (i.e. group isolates)	22388286
sam tools view	0.1.8	Converts a SAM file into a BAM file	http://www.htslib.org/
sam tools sort	0.1.8	Creates a sorted copy of a BAM file	http://www.htslib.org/
fq2fa from IDBA-UD	1.1.1	Converts fastq to fasta input in IDBA-UD	

IDBA-UD	1.1.1	*de novo* Assembler	10.1093/bioinformatics/bts174
get_circular_contigs.csh		Provides a fasta file with the sequences of probable circular contigs	Jorgensen, Xu et al. 2014
get_circularization_paired_reads.csh		Provides a fasta file with the sequences of probable circular contigs	Jorgensen, Xu et al. 2014
BLAST+ , megablast	2.2.18	Finds regions of similarity to databases or genomes	20003500
STAR	2.5.0	Maps short sequence reads to a whole genome focusing on the recognition of spliced transcripts	23104886

2.9.5.1 NGS data analysis

The goal of the procedure followed for the analysis of the obtained data was the identification of circular DNA sequences of non-human origin from breast cancer tissue samples. Therefore, the procedure was designed according to this goal and it was the following:

- Quality Filtering Phred score between 20 and 90
- Removal of sequence of adapter at 3´ end AGATCGGAAGAG (cutadapt)
- Removal of sequence of adapter at 5´ end AGATCGGAAGAG (cutadapt)
- Trimming of Artifacts on
- Sorting by Length > 17
- Removal of all reads which contain “N”s
- Reference genome for mapping and quality check: Human genome 38

- No removal of duplicates in final BAM file
- Tophat2 as aligner for genomic mapping with default parameters, using the EvalRseq software.

The unmapped reads were assembled *de novo* using IDBA-UD with the following parameters:

- minimum contigs 100
- number of threads 5
- maximum kmer 80

The tools used to discover circular contigs from reads (see Table 2.7) were adapted from Jorgensen et al. (Jorgensen, Xu et al. 2014). The corresponding Pearl Scripts were implemented into two shell scripts: get_circular_contigs.csh, and get_circularization_paired_reads.csh (150 bases at contigs ends). Those scripts were run using the assembled contigs from as input. The resulting fasta file contained the sequences of potential circular contigs. The resulting candidates for circular DNA sequences were then cleaned for sequences with more than 50% identity to human DNA by blasting them against the human genome hg38 using the MegaBlast (Blast+, NCBI, parameters -e 0.0001 -m 8 -v 3 -b 3 -D 3 -a 2). Small contigs < 200 bases were removed. In order to detect circular DNAs with similarity to known viruses, finally, a MegaBlast search of the filtered contigs against the database of previously isolated virus genomes was performed.

In a second approach, the mapping of the NGS reads to the human genome 38 was performed using the STAR mapping algorithm (parameters --outReadsUnmapped Fastx --outFilterMultimapNmax 100000 --seedMultimapNmax 500000 --limitOutSJoneRead 100000 --limitOutSAMoneReadBytes 6000000 --outSAMtype BAM Unsorted). The following *de novo* assembly was done as mentioned above. MegaBlast searches against the NCBI non-redundant database (October 2017, 44,394,022 sequences; parameters -v 10 -b 10 -e 0.01) and the viral and

bacterial EMBL databases (September 2017; parameters -v 10 -b 10 -e 0.01) were conducted in order to detect similarities with the circular contig candidates.

2.10 Cell Culture

2.10.1 Cultivation of human cells

Two human cell lines, the L-1236 and L-428, were used in this study. These suspension cells were cultured in suspension culture flasks (Greiner bio-one) in RPMI-1640 Medium, with L-Glutamine and sodium bicarbonate (Sigma Aldrich) containing 10% (v/v) fetal bovine serum (FCS) and 1% (v/v) Pen/Strep (Sigma Aldrich) at 37^{o}C and 5% CO_2. Cells were passaged regularly, maintaining them at a desity of about 0.5-1 x 10^6 cells/ml in order to keep them in an exponential growth phase. During passaging, cells were once washed with sterile 1x PBS.

2.10.2 Mycoplasma test

The cells were regularly tested for contamination with mycoplasma using the Venor®GeM Classic kit (Minerva Biolabs) according to manufacturer's protocol for Taq polymerase 5 U /µl (AmpliTaq Gold DNA Polymerase with Gold Buffer and $MgCl_2$, Applied Biosystems). The presence of mycoplasma-indicating PCR products was assessed by agarose gel electrophoresis on a 1.5% agarose gel (see 2.7.5).

2.10.3 Cell counting

Cell counting was performed with 20 µl cell solution in a Neubauer chamber using a 1:2 dilution of Trypan Blue (Thermo Scientific). The concentration was calculated based on the formula:

$$Concentration = \frac{Number\ of\ cells\ x\ 10000\ x\ 2}{Number\ of\ big\ squares}$$

where 2 corrects for the 1 : 2 dilution of the Trypan Blue and the number of big squares is 4.

In addition to the above mentioned manual cell counting, also an automated cell counting was used in this study. The corresponding Countess™ Cell Counting Chamber Slides (Invitrogen) and the Countess® Automated Cell Counter (Invitrogen) have been used according to the manufacturer's instructions.

2.10.4 Cultivation of new cells

In order to avoid phenotypic changes of the used cell lines due to a high number of passages, fresh vials of the corresponding cell line were taken in culture before each experiment. Cell stocks stored in freezing medium in cryo tubes in liquid nitrogen were thawed rapidly and mixed with fresh RPMI-1640 medium containing 10% FCS and 1% Pen/Strep, pre-heated to 37^{o}C. Cells were then centrifuged at 300 x g for 10 min at room temperature. After washing of the cell pellet with sterile 1x PBS, cells were resuspended in culture medium and transferred into a cultivation flask for further cultivation.

2.10.5 Preparation of frozen cell stocks

Once freshly cultivated cells reached 70-80% confluency, aliquots were stored in liquid nitrogen. Therefore, the cells were centrifuged for 10 min in 300 x g at room temperature and the cell pellet was resuspended in freezing medium at a density of 1 x 10^6 cells per ml. The vials were stored for two days in a box at -80 ^{o}C to assure slow temperature decrease and finally transferred into liquid nitrogen.

2.10.6 DNA transfection

Both cell lines used in this study were transfected with the Nucleofector™ 2b Device (LONZA) using the Nucleofector® Kit V (LONZA) according to the

manufacturer's protocol. Briefly, the required amount of cells (1 x 10^6 cells per sample) were centrifuged at 300 x g for 10 min at room temperature and the supernatant was completely discarded. The cell pellet was dissolved in 100 µl of Nucleofector® solution at room temperature containing the supplement and the desired amount of circular DNA (up to 2 µg) was added. The cell/DNA suspension was transferred into a certified cuvette, which was then placed into the Nucleofector™ 2b Device (LONZA) and the appropriate program for each cell line was applied (X-001 for the L-1236 cells and X-005 for the L-428 cells). Once the program was finished, the cuvette was taken out of the holder and ~500 µl of pre-equillibrated cell culture medium (RPMI-1640 medium containing 10% FCS and 1% Pen/Strep) was added. The mixture was transferred gently into a T-25 cell culture flask containing 4 ml of pre-equilibrated culture medium. Five nucleofection reactions were performed per condition, resulting in a final volume of 7 ml per T-25 cell culture flask. Due to the low proliferation of the L-1236 cells, for these cells two flasks per condition were transfected. A mock transfection using 1x TE buffer instead of DNA was used as a transfection negative control. The transfection efficiency was assessed by counting of green-fluorescent cells 24h and 48 h after the transfection of 2 µg of the pmaxGFP® Vector (LONZA).

2.10.7 Cell harvesting

Harvesting was performed at different time points after transfection. More specifically, total DNA, RNA and protein was harvested at days 3, 7, 10, 14, 28, 35 and 42 post transfection. The cell culture supernatants were collected as well for each time point and stored at -80 ^{o}C. After centrifugation at 300 x g for 10 min at room temperature, the cell pellet was washed once with sterile 1x PBS, before being dissolved in 5 ml fresh culture medium RPMI-1640 medium containing 10% FCS and 1% Pen/Strep. 1 ml of this cell solution was used for harvesting and the remaining 2 ml were transferred back to the flask containing 5 ml of fresh culture medium for further cultivation.

Total DNA was extracted using the DNeasy Blood & Tissue kit (QIAGEN) according to the manufacturer's protocol, while an additional washing step with 1x PBS was included at the beginning. The elution volume was 100 µl. 4 µg if the extracted total DNA were digested with 100 U DpnI enzyme (Fermentas) in a total volume of 400 µl overnight at 37^{o}C in order to remove transfected input DNA. The DpnI-resistant DNA was subsequently precipitated with ethanol (see 2.7.1) and 200 ng were used for replication analyses with PCR or qPCR reactions.

The QIAshredder cell-lysate homogenizers (QIAGEN) were used according to the manufacturer's protocol for lysate homogenization which was stored at -80 ^{o}C for RNA isolation in the future.

For protein extraction, cells were washed with 1 x PBS and subsequently dissolved in 2.5x Lämmli buffer. The mixture was then boiled at 95 ^{o}C for 5 min, followed by 5 min centrifugation at 14000 x g and stored at -20 ^{o}C for further analysis.

2.10.8 Infection of fresh cells with supernatants collected during transfection

Infection experiments of freshly clultured cells were performed using the supernatants collected at the harvesting time points of transfection experiments. Prior to infection, supernatants were centrifuged at 5000 x g for 15 min at room temperature and the supernatant was filtered through 0.22 µm filters (Low protein Binding Durapore, Millipore Millex-GV) in order to remove remaining cells and cellular debris. Three days prior to infection 1.2 x 10^{6} cells were treated with 50 µl 10 mM N-glycolylneuraminic acid (Neu5Gc, Sigma Aldrich) in a total volume of 5 ml cell culture medium (100 µM final concentration). A flask with water instead of Neu5Gc was included as a negative control of each condition. The infected cells were incubated for 2h at 37 ^{o}C and 5% CO_2. Then 3.6 ml of fresh medium RPMI-1640 medium containing 10% FCS and 1% Pen/Strep and 50 µl 10 mM Neu5Gc (Sigma Aldrich) or H_2O were added and incubation at the above-mentioned conditions

was continued. Total DNA was harvested at days 3, 7, 10 and 14 post infection from 2 ml cell solution. During each passaging step, Neu5Gc was added at a final concentration of 100 µM.

2.10.9 Phenotypic analysis

2.10.9.1 Cell viability assay

In order to assess cell viability, a cell viability assay was performed in a 96-well-plate format using the CellTiter-Glo Luminescent Cell Viability Assay kit (Promega) using 50000 cells in a final volume of 100 µl in each well. The luminescent signal was measured after 24h, 48h and 72h according to the manufacturer's protocol and cell viability was calculated according to it.

2.10.9.2 Cell proliferation assay

Cell proliferation was tested using the Click-iT^{TM} Edu Plus Alexa Fluor^{TM} 647 Imaging Kit (Thermo Scientific) as recommended by the manufacturer. Cell proliferation was measured after 24h, 48h and 72h following the manufacturer's protocol. The cells were fixed by a drop wise addition of cold 70% ethanol before being washed with 5% FCS in 1x PBS. The cells were analyzed by flow cytometry using 633 / 635 nm excitation with a red emission filter.

2.11 Protein biochemical methods

2.11.1 SDS PAGE

Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS PAGE) was used for separation of proteins according to their molecular weight. SDS is a detergent with a substantial protein-denaturing effect, that along with boiling and a reducing agent (such as β-mercaptoethanol), disrupts the tertiary structure of proteins unfolding them into linear chains with negative charge

proportional to the polypeptide chain length. Therefore, the proteins are separated solely by their size in the polyacrylamide gel in an applied electrical field. For protein separation the commercially available TruPAGE™ Precast Gels 4-20% (Sigma Aldrich) were used. The protein solutions were pre-heated at 95 ^{o}C and 15 µl were loaded on the gel. 5 µl Page Ruler™ Prestained Protein Ladder (Fermentas) (**Fig. 2.2**) were loaded, as well as 5 µl positive control (MSBI1.176 codon optimized Rep protein). Gels were run at 100 mA in 1x TruPAGE SDS running buffer (Sigma Aldrich).

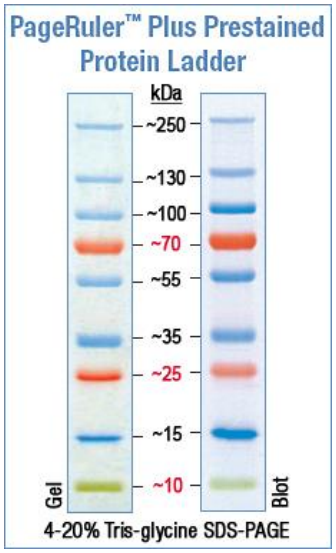


Figure 2.2 Page Ruler™ Prestained Protein Ladder (Fermentas).

2.11.2 Immunoblotting

The separated proteins were transferred to a nitrocellulose membrane using the Trans-Blot® Turbo™ Transfer System (Bio-Rad) according to the manufacturer's instructions using the MIXED MW protocol.

2.11.3 Ponceau S staining

Ponceau S was used for rapid detection of microgram quantities of proteins transferred to a nitrocellulose membrane. Ponceau S staining is a reversible

technique that allows further immunological detection. For staining, the membrane was incubated on a shaker in a sufficient amount of Ponceau S staining solution. Afterwards, it was washed two times with water and the protein-less background was defined. The membrane was completely destained by washing with water containing 8 mM NaOH for 10 min. A additional washing step with water was performed in order to remove remaining NaOH that would interfere with the downstream processes.

2.11.4 Immunodetection

Immunological detection was used to identify specific proteins after blotting. This is a chemiluminescence-based method, using a horseradish peroxidase (HRP)-conjugated secondary antibody and an enzymatically activated substrate for the detection of the primary antibody which is specific for the protein of interest. Following transfer, blocking is an important step that prevents antibodies from binding to the membrane non-specifically. The membrane was incubated in blocking buffer for 1 h on a shaker at room temperature followed by addition of the primary antibody (1:5000 dilution) and an overnight incubation at 4 ^{o}C. The blot was washed four times for 10 min in washing buffer before the secondary antibody was applied in 5% skim milk / PBS buffer at a dilution of 1:5000 for 1h on a shaker at room temperature. Finally, four washing steps in washing buffer, for 10 min each, were performed. Then, the blot was developed using the Clarity^{TM} Western ECL Substrate (Bio-Rad) in a 1:1 mixture of peroxide and lumino/enhancer solution. Development was performed in a blot imager. The membranes were either stored or treated with additional antibodies.

2.11.5 Stripping of membranes

In certain cases, one membrane was probed with two different primary antibodies. Therefore, the previous primary antibody had to be efficiently removed by stripping. Therefore, the membrane wasincubated for 15 min at

37 ^{o}C in Restore^{TM} PLUS Western Blot Stripping Buffer (Thermo Scientific) followed by two washing steps with washing buffer , before incubation with the new primary antibody of interest.

2.12 ELISA experiments & serology

Serum samples from breast cancer and Hodgkin's Disease patients were analyzed for human anti-MSBI1.176 Rep antibodies using ELISAs.

2.12.1 ELISA protocol

200 ng per well of antigen were diluted in PBS 4M Urea and 100 µl of diluted antibody were pipetted in each well of microtiter plate (MaxiSorp, Thermo Scientific). Each sample was in triplicates. The plate was covered with adhesive plastic and incubated at 4^{o}C overnight. The coating solution was removed and the plate was washed with 150 µl PBS. Blocking was done with 100 µl Superblock® T20 (PBS) Buffer (Thermo Scientific) and incubation for 2 h at room temperature on a shaker. Subsequently a 1:500 dilution of serum samples was prepared in Superblock and 100 µl were added in each well and incubated for 2 h in 37^{o}C. The blocking solution was then removed and the plate was washed three times with 150 µl PBS 0.1% Tween 20, with the first washing to be directly removed, while for the two remaining a 10 min incubation was included. Washing was followed by 1 h incubation of a HRP-coupled goat anti-human secondary antibody (100 µl in each well) in a dilution of 1:5000 in Superblock on a shaker in room temperature. The plate was subsequently washed 3x with 150 µl PBS 0.1% Tween 20. The color development was done by adding 100 µl TMB Substrate Solution (Thermo Scientific) and stopped by adding 100 µl stop solution. Absorbance was recorded at 450 nm.

3. Results

3.1 Isolation and identification of episomal DNA agents from cow milk and dairy products

A number of circular episomal DNA agents has been isolated by our group from commercially available cow milk and healthy bovine cattle, named bovine milk and meat factors (BMMFs), as well as from blood and tissue samples from MS patients (Funk, Gunst et al. 2014, Gunst, Zur Hausen et al. 2014, Lamberto, Gunst et al. 2014, Whitley, Gunst et al. 2014). Part of this thesis was to continue this work and to expand it to different types of dairy products. Therefore, DNA was isolated by phenol-isoamyl-chloroform extraction from commercially available cow milk, yogurt, card cheese, crème fraiche and butter (see Appendix for a full list of the products). Subsequently, the DNA preparations were subjected to Rolling Circle Amplification (RCA) using random primers, in order to specifically amplify small circular DNA molecules. Specific primers (see 2.3) for the BMMF1 group were used in order to identify such agents related to the already identified DNA agents (zur Hausen, Bund et al. 2017) by conventional PCR (**Figure 3.1**). The obtained DNA molecules were subsequently cloned and sequenced.

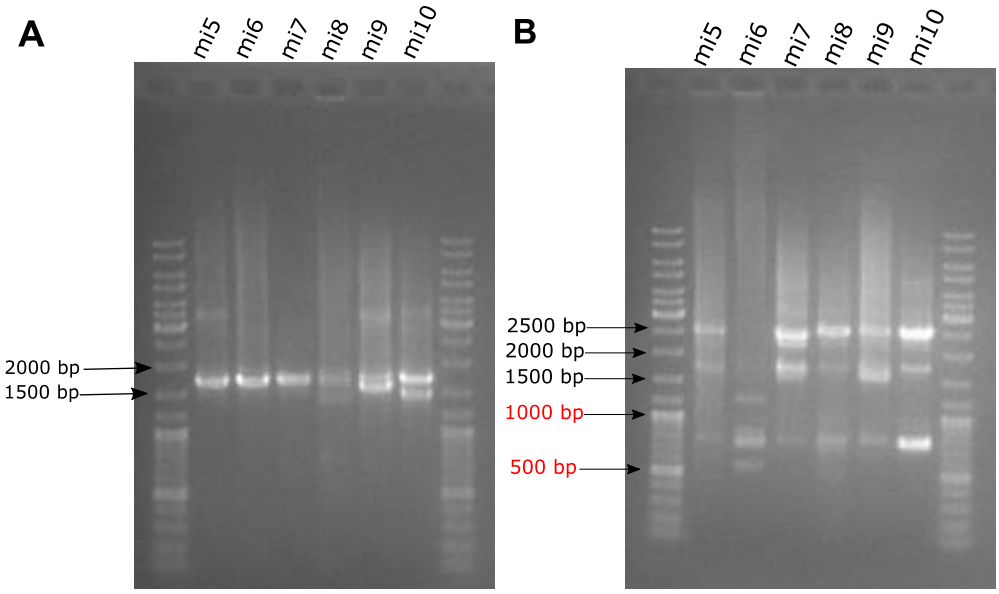


Figure 3.1 Long-PCR with specific primers on DNA isolated from commercially available yogurt and milk samples after RCA. **A)** Bands obtained with the NnXn primers and **B)** bands obtained with the NoXo primers. The obtained molecules were cloned and sequenced.

By this approach, two novel circular DNA sequences were isolated from the same milk sample (mi7, more information in Appendix), CMI5.170 (1706 bp) and CMI5.240 (2406 bp) (Falida, Eilebrecht et al. 2017) (**Figure 3.2**).

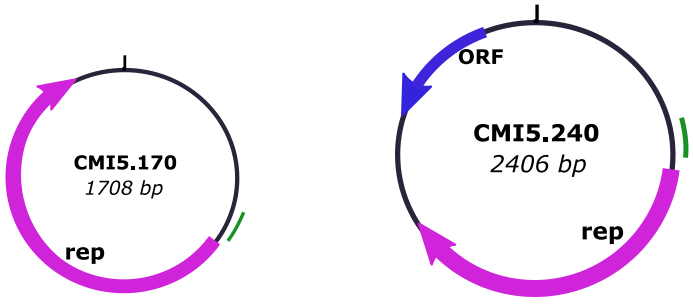


Figure 3.2 Circular representation of the newly isolated agents CMI5.170 and CMI5.240. The Rep protein is represented in pink, the additional open reading frame of the CMI5.240 is indicated in blue and the green line indicates the conserved repeat region upstream of the Rep open reading frame.

CMI5.170 shares 89% nucleotide identity to the previously identified MS brain isolate MSBI1.170 (1766 bp), while for CMI5.240 the nucleotide identity to MSBI1.176 is 68%. **Figure 3.3** summarizes the similarities among the isolates of the BMMF1 group including the newly isolated genomes CMI5.170 and CMI5.240.

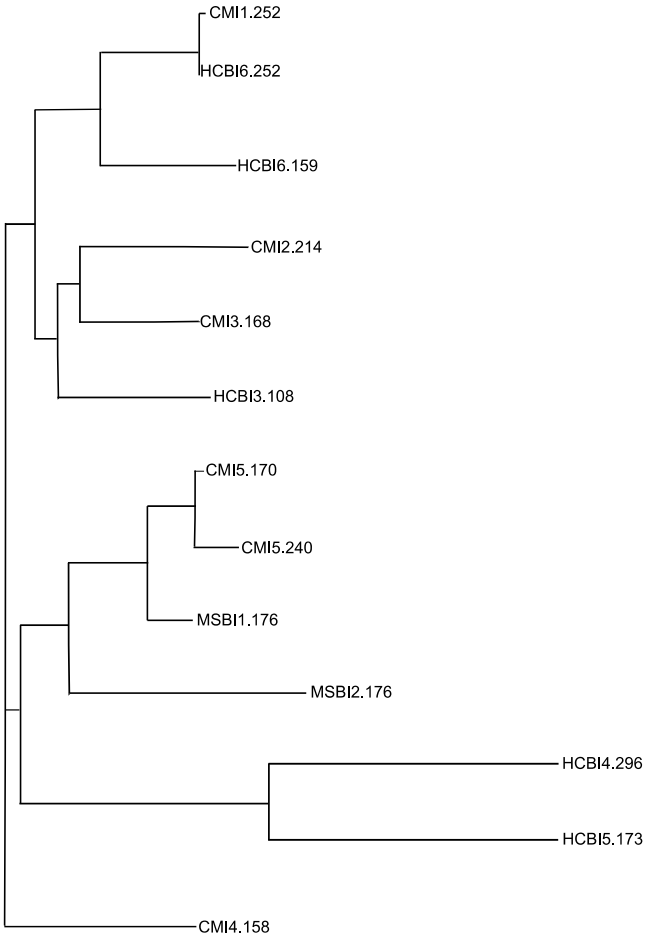


Figure 3.3 ClustAL W dendogram showing the similarity of the BMMF1 group and the newly isolated CMI5.170 and CMI5.240.

The open reading frame (ORF) encoding a putative replication associated protein is highly conserved among all BMMF1 isolates. The encoded putative Rep proteins show identities of 50-98% of amino acids. CMI5.240 contains an additional ORF of 125 aa in the antisense direction and it is 96% identical at amino acid level to an antisense-directed ORF found in

CMI2.214 (Whitley, Gunst et al. 2014). Both isolates contain a highly conserved A/T-rich region, which is followed by a repeat region upstream of the Rep ORF as it has been previously reported for the other isolates (zur Hausen, Bund et al. 2017) (**Figure 3.4**).

Apart from these new isolates, three of the previously isolated agents were also isolated from additional dairy products. CMI3.168 was isolated from milk, yogurt and curd cheese, CMI4.158 from milk and HCBI6.252 from milk and yogurt (Falida, Eilebrecht et al. 2017).

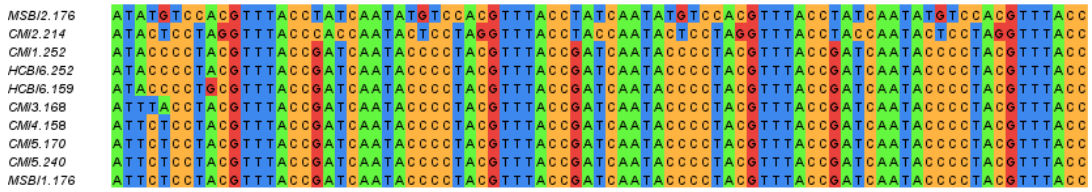


Figure 3.4 CLUSTAL alignment of the repeat region of the BMMF1 group and the newly isolated CMI5.170 and CMI5.240.

Specific back-to-back primer sets located within the conserved region of BMMF2 group of isolates (Sphinx2.36-related) have been used for long-PCR reactions using DNA samples isolated from the above mentioned dairy products and additional milk samples as a template, revealing a number of additional, novel isolates (15-30 new isolates) and indicating that there is a significant multitude of such circular episomal agents that might have a potential role in human diseases.

3.2 Identification of episomal DNA agents in breast cancer biopsies by Next Generation Sequencing (NGS).

In addition to the classical molecular biology-based DNA isolation techniques shown before, also isolation of DNA from breast cancer tissue samples by NGS was performed. Milk and dairy product consumption has been previously linked to breast cancer (zur Hausen and de Villiers 2015). Assuming a role of BMMF infection during breast cancer formation, the corresponding BMMF genomes should be identifiable by NGS of total DNA isolated from breast cancer tissue, since a transfer of BMMFs to their target cells is assumed after initial infection early in life. For the identification of novel agents, a high throughout NGS approach has been chosen, since it represents a powerful tool and the obtained data can be subjects to metagenome based analysis for the discovery of novel human viruses (Barzon, Lavezzo et al. 2011). This requires special sample preparation and enrichment in order to detect very low abundance of virus populations over the cellular DNA background.

Therefore a protocol was established specifically for the isolation of circular DNA molecules by comparing four different DNA extraction methods (see 2.8.1). In order to test the efficacy of each method, known amounts of plasmid vector (pUC19, 2700 bp) were added to approximately the same amount of placenta tissue (25 mg) before DNA isolation. This spiked tissue was then used as testing material. Four different vector amounts were used, representing approximately 10 pg, 2 pg, 200 fg and 100 fg.

After DNA extraction with the four methods, long-PCR was performed using specific primers for pUC19 (**Figure 3.5**).

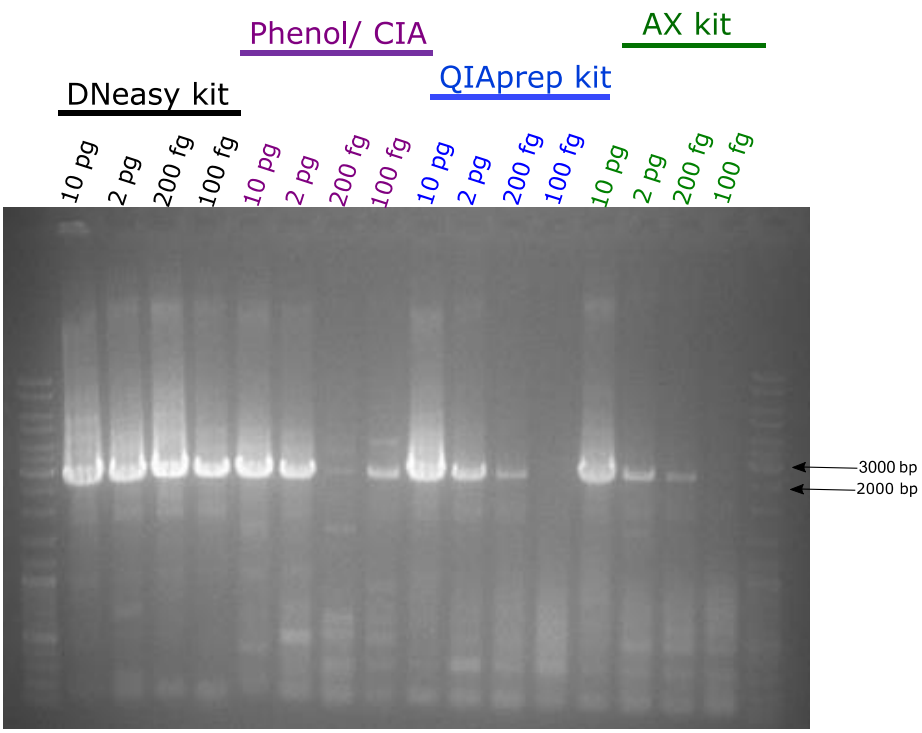


Figure 3.5 Long-PCR on DNA extracted from placenta tissue with indicated amounts of pUC19 plasmid using pUC19-specific primers and comparing different sample preparation techniques. The band intensity reduces as the amount of pUC19 reduces for the plasmid kits, while it remains stable for the DNeasy kit. The observed difference between the conditions of 200 fg and 100 fg of pUC19 in the phenol/chloroform extraction was due to technical problems in the long-PCR.

The samples with the lowest amount of pUC19 DNA (3 molecules of pUC19 / 100 cells) were used for further processing. Additionally, a larger plasmid was used in a parallel approach (Nanoluc vector, 5900 bp) in order to test for differences on the enrichment between small and large DNA molecules (see below).

RCA was used for a specific enrichment of circular DNA molecules. In addition, the effectivity of a Plasmid-Safe DNase treatment was tested, which selectively digests contaminating linear genomic DNA molecules from plasmid DNA preparations. The enrichment of the plasmid DNA after RCA as well as the effectiveness of the Plasmid-Safe DNase digestion were tested by RT-

qPCR analyses for both plasmids (pUC19, **Figure 3.6** and Nanoluc, **Figure 3.7**). A significant enrichment of the plasmid DNAs after RCA was observed in all extraction methods. Notably, an increase of up to 350 fold was observed after Plasmid-Safe DNase digestion.

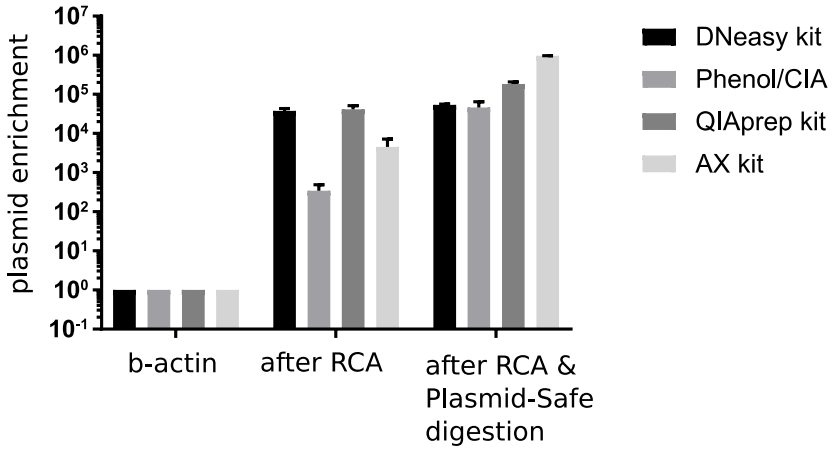


Figure 3.6 RT-qPCR data representing the enrichment of pUC19 plasmid DNA in each extraction method before and after RCA as well as after RCA upon Plasmid-Safe DNase digestion. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

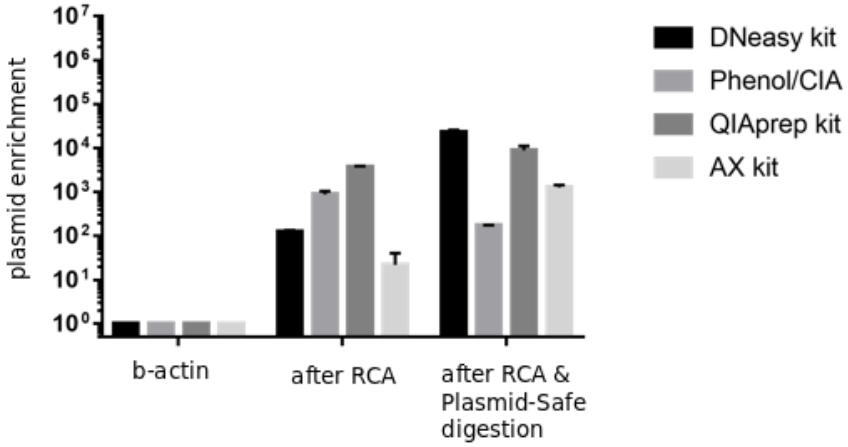


Figure 3.7 RT-qPCR data representing the enrichment of Nanoluc plasmid DNA in each extraction method before and after RCA as well as after RCA

upon Plasmid-Safe DNase digestion. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

The DNA isolation with the DNeasy kit (QIAGEN) showed the strongest enrichment for larger plasmids such as the Nanoluc vector. DNA extraction using the Plasmid Mini AX kit (A&A Biotechnology) most efficiently enriched smaller plasmids such as the pUC19 vector. Therefore, these two kits were chosen for isolation of DNA from the breast tissue samples.

Eleven breast tumor tissue samples and their respective tissue adjacent to tumor (TAT) control tissues of the same patients were used for DNA extraction. DNA yields in tumor tissue were significantly higher than in normal tissues, as normal breast control tissue is predominantly fatty tissue containing significantly less cells than tumor tissue. As it was not possible to obtain sufficient amounts of DNA for further processing using the Plasmid Mini AX kit (A&A Biotechnology), all further analyses were performed with the total DNA obtained with the DNeasy kit (QIAGEN). The isolated DNA was subsequently subjected to Plasmid-Safe DNase digestion followed by RCA with random primers and ethanol precipitation.

One patient sample (tumor and control) was used for a NGS pilot experiment. In this pilot experiment, the vast majority of potentially circular contigs were identified as human sequences, necessitating a more efficient removal of human reads before contig generation. Moreover, the samples were initially amplified by RCA and thus differ substantially from typical genomic DNA samples, which may have negatively impaired sequencing depth. For example, the DNA fragments used for library generation were larger than the optimal ones resulting in a significantly reduced number of reads. Therefore, the library preparation method was changed to a protocol for formalin-fixed paraffin-embedded (FFPE) samples for the remaining samples.

This protocol for library generation was successfully performed for DNA from seven tumor and their corresponding TAT samples. The sequencing was performed using a HiSeq4000 Illumina platform generating paired-ends of 100 bp. The samples were pooled in one lane using different barcodes in order to be able to distinguish the different samples during data analysis and each sample had an adequate number of reads obtained after sequencing (**Table S3**, in Appendix).

After generation of potentially circular contigs, these were analyzed for similarities to the previously identified BMMFs. One circular contig of 1620 bp size was identified in a TAT sample (7717N), which showed 85% similarity to the previously isolated MSSI.162 (Gunst, Zur Hausen et al. 2014) according to BLAST search (**Figure 3.8,** 7717N_contig_80). The significant differences between the two genomes at the nucleotide level exclude the possibility of a laboratory contamination. This contig showed also 95% similarity to an uncultured bacterium plasmid clone HD10Abpcirc putative replication protein gene isolated from plasma of patients in maintenance hemodialysis deposited in the GenBank under the accession number KX838914 (Biagini,P.).

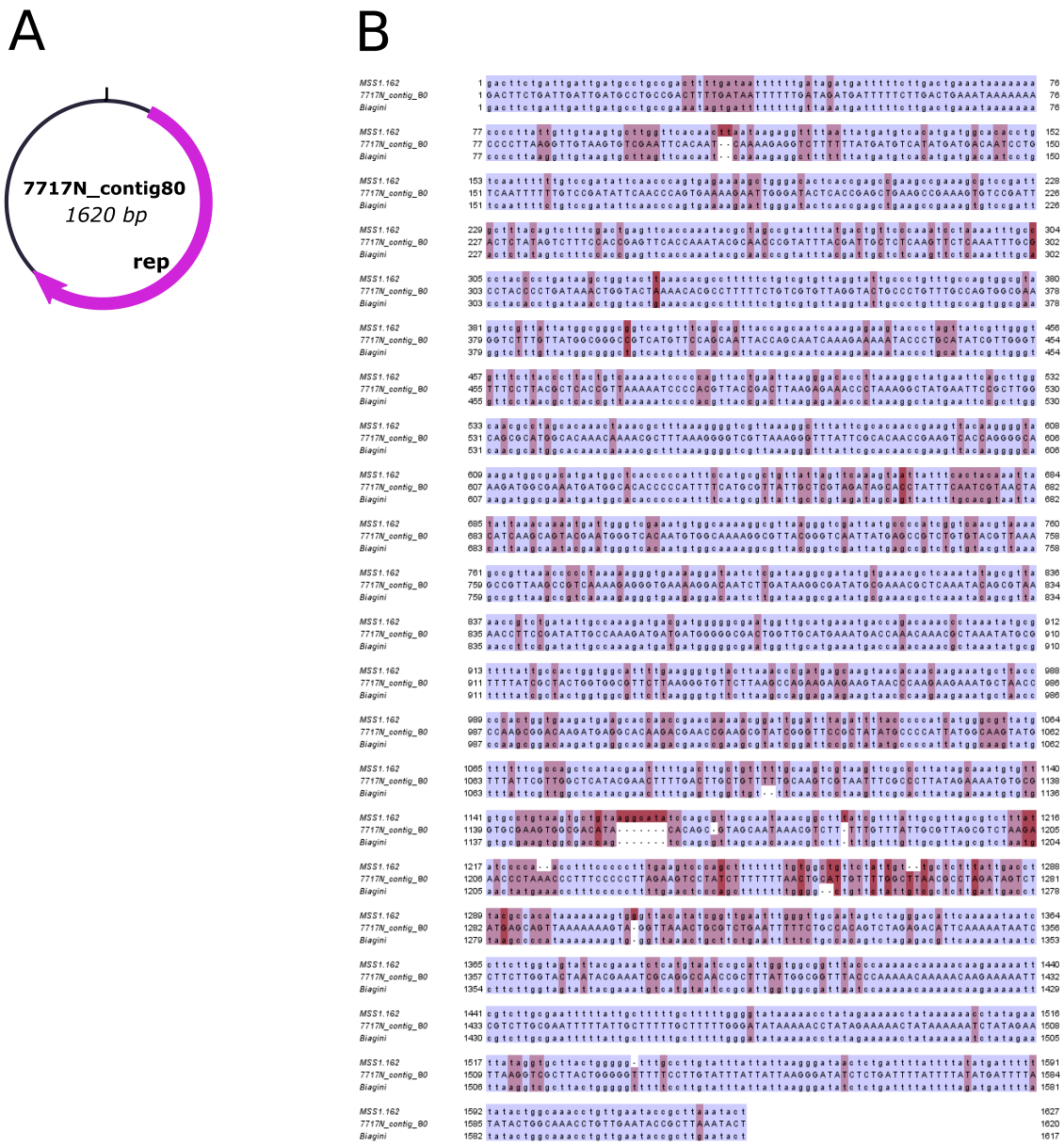


Figure 3.8 A) Circular representation of the contig found in breast TAT tissue sample. **B)** Alignment of the circular contig with the previously isolated MSSI1.162 and the plasmid gene isolated from Biagini (KX838914).

Back-to-back primers were designed based on the 7717N_contig_80 sequence and PCR reactions using these primers were performed in all fourteen initial RCA samples used in this study. Solely the 7717N sample showed a positive signal at the expected height, proving the physical presence of the corresponding circular DNA molecule (**Figure 3.9**)

One of the samples (7728T) was excluded from further analysis due to contamination with circular MSBI1.176 genome cloned to pUC19 vector as revealed from the analysis. Although this was clearly a contamination, it was at the same time a confirmation of the success of the analytical process that was used.

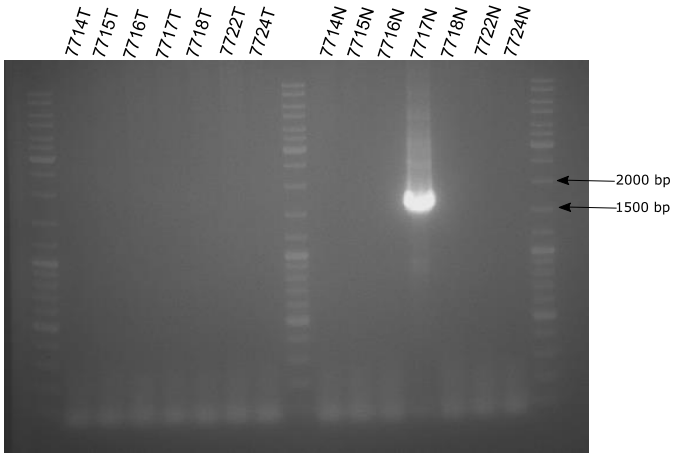


Figure 3.9 PCR with specific back-to-back primers on the seven breast tumor tissue samples (7714T-77124T) and the corresponding TAT controls (7714N-7724N) confirming the presence of the new isolate only in the expected (7717N) sample with approximately 1600 bp size.

In a parallel effort, restriction digestion analyses using various restriction enzymes and PCR analyses using specific back-to-back primers for the previously identified agents were performed on the initial RCA samples. However, none of these approaches resulted in the identification of novel episomal DNA agents. All resulting DNA fragments cut and cloned corresponded to human DNA sequences.

3.3 Replication of Cattle-Derived Milk and Meat factors (BMMF) in human cells and analysis of their involvement in human breast cancer and in Hodgkin´s Lymphomas.

This project focused on the search for a permissive cell system for the identified BMMFs. Several epidemiological studies point to a link between the consumption of milk and red meat and general animal exposure and the development of Hodgkin's disease (Khuder, Mutgi et al. 1999, Epstein, Chang et al. 2015). In addition, 24 full-length TT Viruses (a single-stranded DNA virus) have been isolated from a single spleen biopsy sample from a patient with Hodgkin's lymphoma (Jelcic, Hotz-Wagenblatt et al. 2004). In order to test replication competence of BMMFs in cells derived from Hodgkin's disease, two Hodgkin's Lymphoma cell lines, the L-1236 and L-428 cell lines, were selected for transfection with the Multiple Sclerosis brain isolate MSBI1.176, which was recently shown to have the highest transcriptional activity and protein expression in human HEK293TT cells (Eilebrecht, Hotz-Wagenblatt et al. 2018). The L-1236 and the L-428 cells are pre-B cells (Kuppers, Rajewsky et al. 1994). Blood cells of the B cell lineage may probably transport BMMFs to the affected organ (in this case lymphoid tissue). Notably, previous experiments (zur Hausen, unpublished data) showed small putative particles in the L-1236 cells after infection with herpes simplex virus (HSV-1), which may hint at viral infections playing a role in the pathogenesis of Hodgkin's disease.

3.3.1 Transfection with MSBI1.176

To analyze replication of BMMFs in the above mentioned cell lines, 5-10×10^6 cells were transfected with circularized MSBI1.176 genome and a mock condition with addition of TE buffer instead of DNA was included as a negative control. Total DNA and protein were collected at days 3, 7, 10, 14, 28 and 36 post transfection and supernatants were collected and frozen at -

80^{o}C (see 2.9.6 and 2.9.7). DNA was digested with DpnI restriction enzyme in order to remove methylated input DNA while retaining non-methylated replicated DNA. Subsequently, PCR reactions were performed with the DpnI–digested samples as a template using specific primers for the detection of MSBI1.176 (**Figures 3.10A, 3.11A)**.

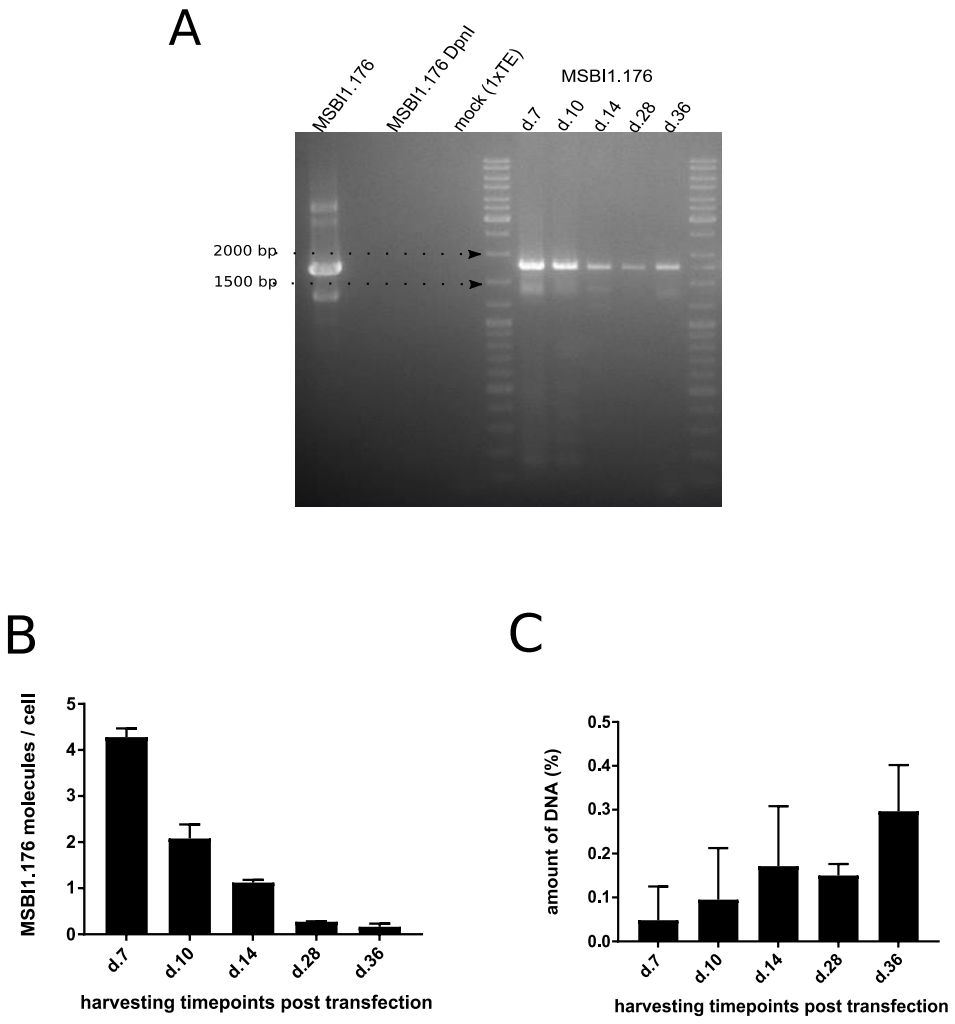


Figure 3.10 A) PCR after DpnI digestion on DNA extracted from L-1236 cells after transfection with MSBI1.176 using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction and DpnI-digested circular MSBI1.176 was used as a template for a negative control reaction. **B)** qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-1236 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

C) qPCR data representing the ratio of MSBI1.176 DpnI digested DNA to MSBI1.176 non-DpnI digested DNA at different harvesting time points of L-1236 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

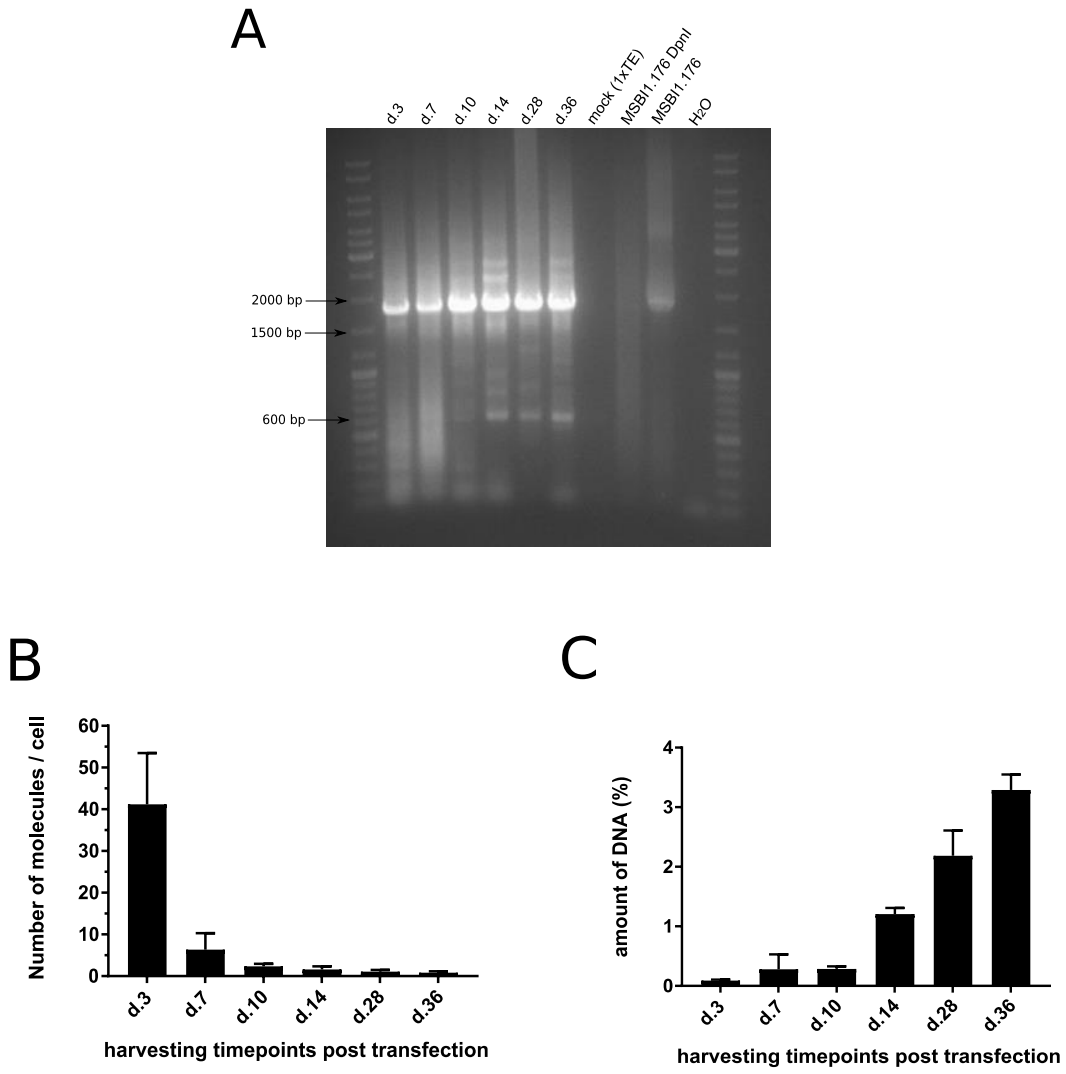


Figure 3.11 A) PCR after DpnI digestion on DNA extracted from L-428 cells after transfection with MSBI1.176 using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction and DpnI-digested circular MSBI1.176 was used as a template for a

negative control reaction. **B)** RT-qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-428 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA. **C)** RT-qPCR data representing the ratio of MSBI1.176 DpnI digested DNA to MSBI1.176 non-DpnI digested DNA at different harvesting time points of L-428 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

MSBI1.176 was detected up to day 36 post-transfection in both cell lines using the DpnI-digested DNA samples as a template for PCR reactions, pointing at synthesis of new, non-methylated DNA molecules during productive replication of MSBI1.176. While the band intensity declines towards later time points in L-1236 cells, it remains relatively stable until day 36 post transfection in both cell lines.

Quantitative-PCR (qPCR) confirms the above mentioned results obtained by PCR. Although the number of MSBI1.176 molecules decreased, it was still possible to detect approximately 1 molecule of MSBI1.176 per cell after 36 days post transfection in L-428 cells (**Figure 3.11B)** and 1 molecule per 5 cells in L-1236 cells (**Figure 3.10B**). The ratio of newly synthesized DNA molecules (DpnI-resistent) to input DNA (DpnI-sensitive) supports a productive replication (**Figures 3.10C, 3.11C)**.

Western Blot analyses using the protein samples extracted at different time points after transfection with MSBI1.176 using antibodies specific for the detection of MSBI1.176 Rep protein did not reveal a signal at the expected protein size (38 kDa) most likely0 due to the low general translation activity of BMMF1 genomes which was characterized in a recent study (Eilebrecht, Hotz-Wagenblatt et al. 2018) (see Figure in Appendix D).

3.3.2 Co-transfection experiments

In previous studies it has been demonstrated that MSBI1.176 replicates abortively in HEK293TT cells (Eilebrecht, Hotz-Wagenblatt et al. 2018). However, co-transfection with CMI1.252 resulted in an increase of MSBI1.176 replication. This detected helper effect aligns with similar phenomena observed in several viruses (Buller, Janik et al. 1981, Rosario, Duffy et al. 2012). Based on these data, co-transfection experiments were performed in the L-428 cell line with different genomes in order to test a helper or interfering effect on the stable replication of MSBI1.176.

3.3.2.1 Co-transfection of L-428 cells with MSBI1.176 and CMI1.252

After co-transfection of L-428 cells with MSBI1.176 and CMI1.176, no effect was observed on the replication of MSBI1.252 after qPCR analysis. As MSBI1.176, CMI1.252 can replicate up to day 28 in this cell line when individually transfected, as well (**Figure 3.12**).

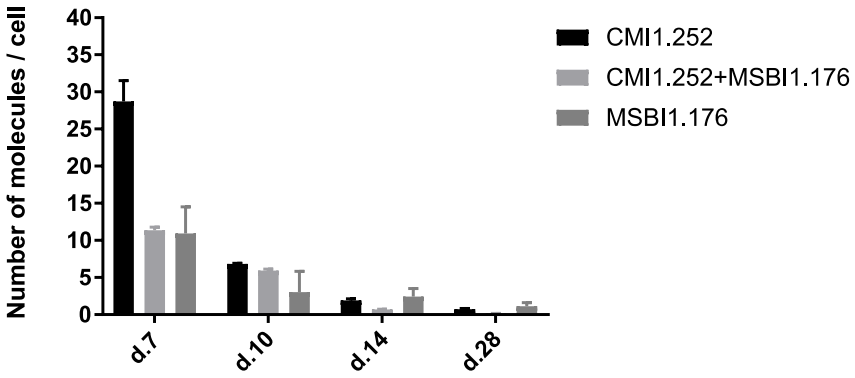


Figure 3.12 RT-qPCR data representing the number of CMI1.252 molecules per cell at different harvesting time points of L-1236 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

3.3.2.2 Co-transfection of L-428 cells with MSBI1.176 and sg613

There was a reproducible observation of extra-bands appearing at different sizes after transfection with MSBI1.176 in replication experiments. Due to its reproductive detection, a band of approximately 600 bp in the L-428 cell line transfections (**Figure 3.11A**) appeared to be of special interest. This genome (named small genome 613, sg613) consisting of the MSBI1.176 region covering the fourth repeat of the repeat region up to approximately $2/3$ of the Rep ORF (analyzed by E.M. de Villiers) was used to transfect the corresponding cell line. The results obtained from long-PCR with specific primers for the detection of the small genome showed that sg613 can replicate independently in this cell line (**Figure 3.13**).

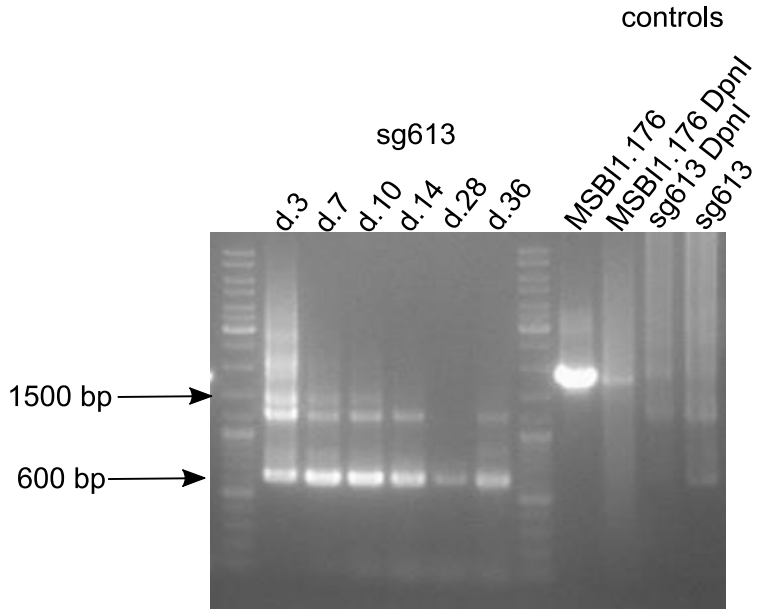


Figure 3.13 PCR after DpnI digestion on DNA extracted from L-428 cells after transfection with sg613 using sg613-specific back-to-back primers. Circular MSBI1.176 and sg613 were used as template for a positive control reaction and DpnI-digested circular MSBI1.176 and sg613 were used as template for a negative control reaction.

Subsequently, this small genome was used for co-transfection experiments with MSBI1.176 in order to explore a possible helper or regulatory effect on the replication of MSBI1.176. However, no modulatory effect was detected on MSBI1.176 replication, which was checked by specific MSBI1.176 detection primers with long-PCR as well as with qPCR (**Figure 3.14**). The amount of newly synthesized DNA molecules to input DNA confirmed that there is no significant effect on the replication of MSBI1.176, although a slightly elevated amount of MSBI1.176 molecules was detected in the single condition compared to the co-transfection.

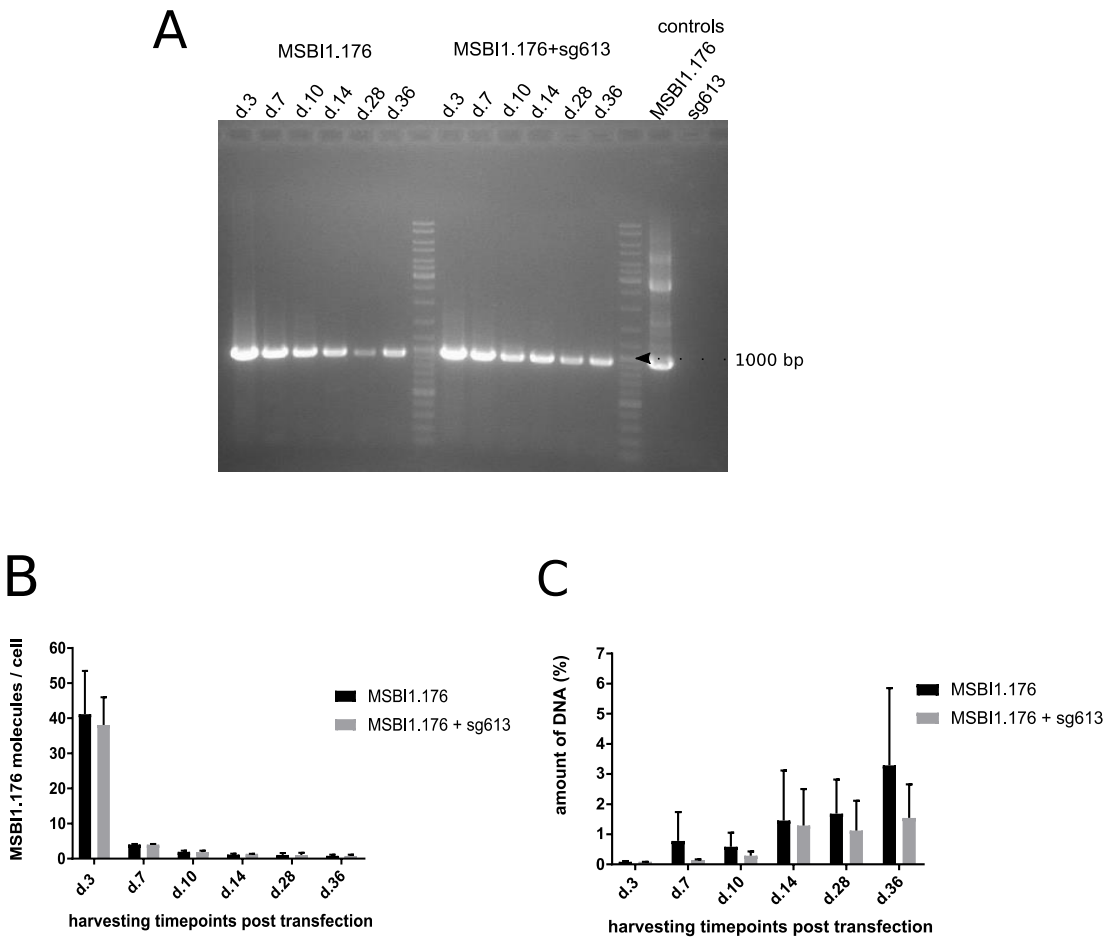


Figure 3.14 A) PCR after DpnI digestion on DNA extracted from L-428 cells after transfection with MSBI1.176 and co-transfection with MSBI1.176 and sg613 using MSBI1.176-specific back-to-back primers. Circular MSBI1.176

and sg613 were used as template for a positive control reaction and DpnI-digested circular MSBI1.176 and sg613 were used as template for a negative control reaction. **B)** RT-qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-428 cells for the single MSBI1.176 condition and the co-transfection with sg613. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA. **C)** RT-qPCR data representing the ratio of MSBI1.176 DpnI digested DNA to MSBI1.176 non-DpnI digested DNA at different harvesting time points of L-428 cells for the single condition of MSBI1.176 and the co-transfection with sg613. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

3.3.3 Transfection with MSBI1.176 and external addition of Neu5Gc

It has been recently reported that the incorporation of N-Glycolylneuraminic acid (Neu5Gc) into human cells leads to chronic inflammation which potentially promotes cancer and other diseases (Alisson-Silva, Kawanishi et al. 2016). Neu5Gc cannot be synthesized in humans due to an exon deletion mutation in the enzyme responsible for the conversion of Neu5Gc from its precursor N-acetylneuraminic acid (Neu5Ac). In humans, Neu5Gc is metabolically incorporated via dietary consumption especially of red meat and cow milk. A number of reasons have been recently summarized supporting the idea that a functional receptor for BMMFs is established by Neu5Gc modification (zur Hausen, Bund et al. 2017).

In order to test an involvement of Neu5Gc in the permissiveness of human cells for BMMF replication, a condition of MSBI1.176 transfection with external addition of 100 µM of Neu5Gc after day 3 post transfection and at every harvesting time point of the L-1236 cell line was included. However, no significant difference was detected after long PCR and qPCR, although the signal is slightly elevated for the condition with the external Neu5Gc during the first time points (**Figure 3.15**).

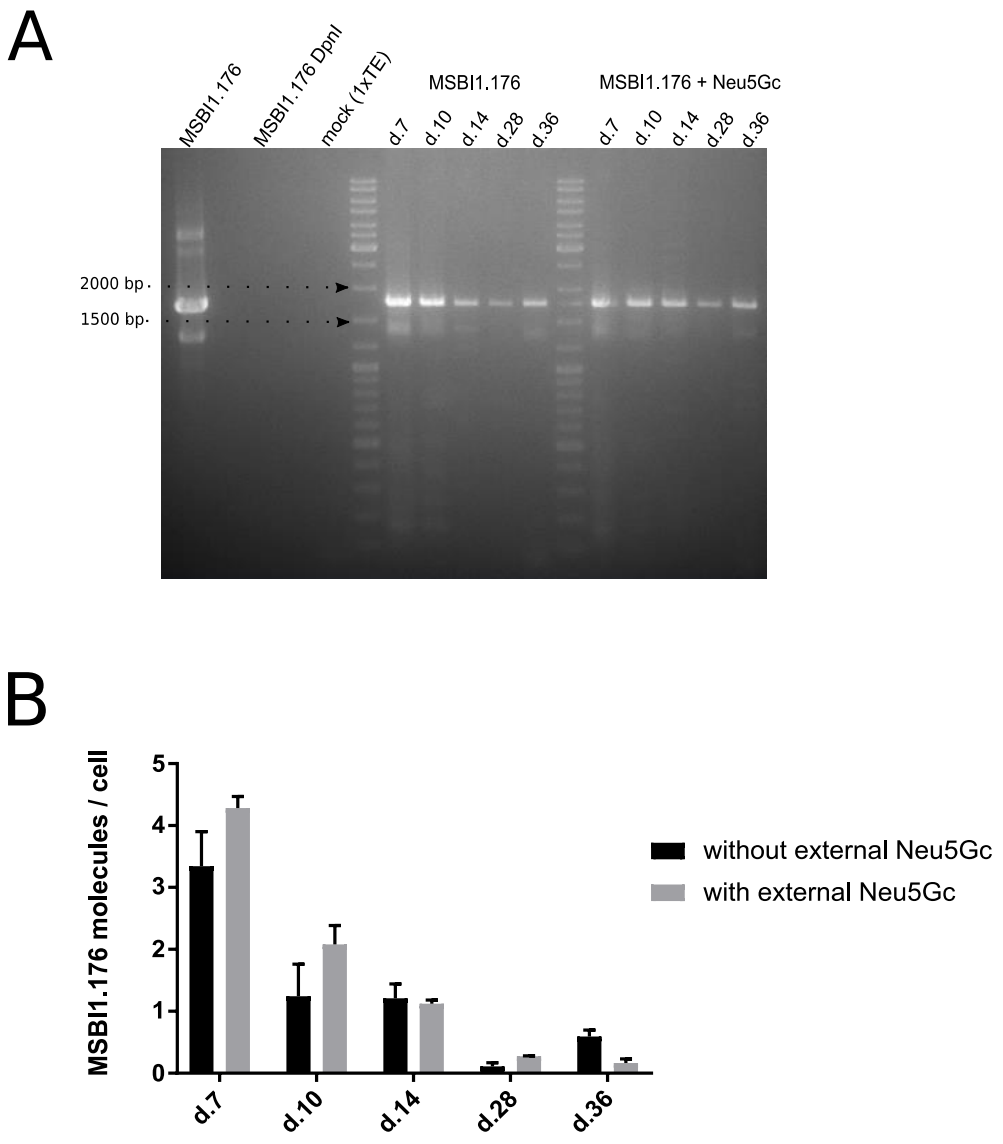


Figure 3.15 A) PCR after DpnI digestion on DNA extracted from L-1236 cells after transfection with MSBI1.176 and MSBI1.176 with external addition of 100 µM Neu5Gc using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction and DpnI-digested circular MSBI1.176 was used as a template for a negative control reaction. **B)** qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-1236 cells with and without addition of external Neu5Gc. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

3.3.4 Infection experiments

As mentioned earlier, Neu5Gc may be involved in the permissiveness of human cells for BMMF infection, In order to establish a general infection system for BMMFs and - in parallel - to test an involvement of Neu5Gc, infection assays were performed using MSBI1.176 positive supernatants of MSBI1.176-transfected cells (**Figure 3.16**).

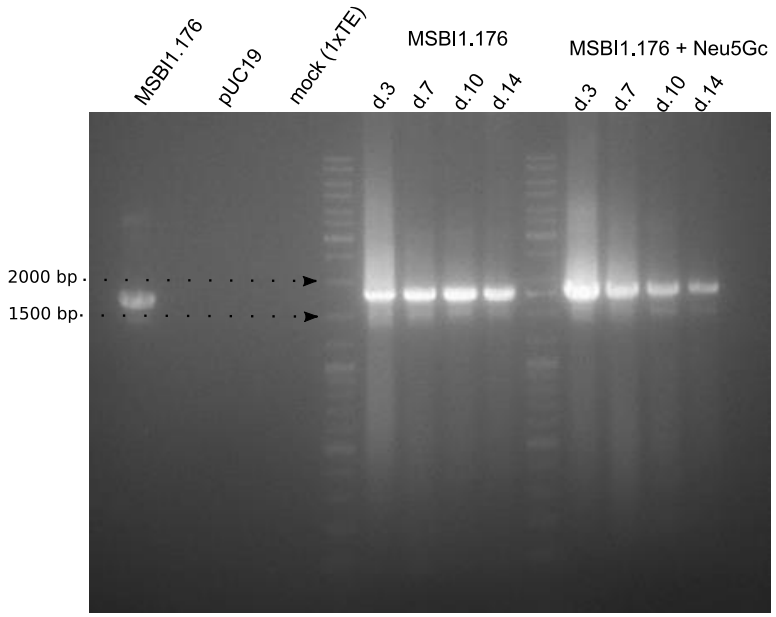


Figure 3.16 PCR analysis of DNA extracted from supernatants collected after transfection of L-1236 cells with MSBI1.176 and MSBI1.176 pre-treated with Neu5Gc after day 3 post transfection and at each harvesting time point using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction. Transfection with pUC19 vector and 1xTE buffer were used as negative control reactions.

Supernatants from cells at day 7 after transfection with i) MSBI1.176, ii) MSBI1.176 upon treatment of cells with Neu5Gc on day 3 after transfection, iii) pUC19 and iv) mock transfection with 1xTE buffer, were used for infection of the L-1236 cell line. Two cell culture conditions were tested, one in the presence of an approximately 100-fold excess of Neu5Gc and one without

addition of external Neu5Gc. Total DNA was harvested at different time points post infection and subsequently analyzed for the presence of MSBI1.176 by PCR using specific back-to-back primers, but no DNA uptake was observed for the target cells (**Figure 3.17**). The same applies for corresponding experiments performed in the L-428 cells (data not shown). However, infectivity of BMMFs remains elusive and technical improvements need to be done in order to further analyze a potential infectious system for BMMFs and prove an activation of the putative cellular receptors by a Neu5Gc modification.

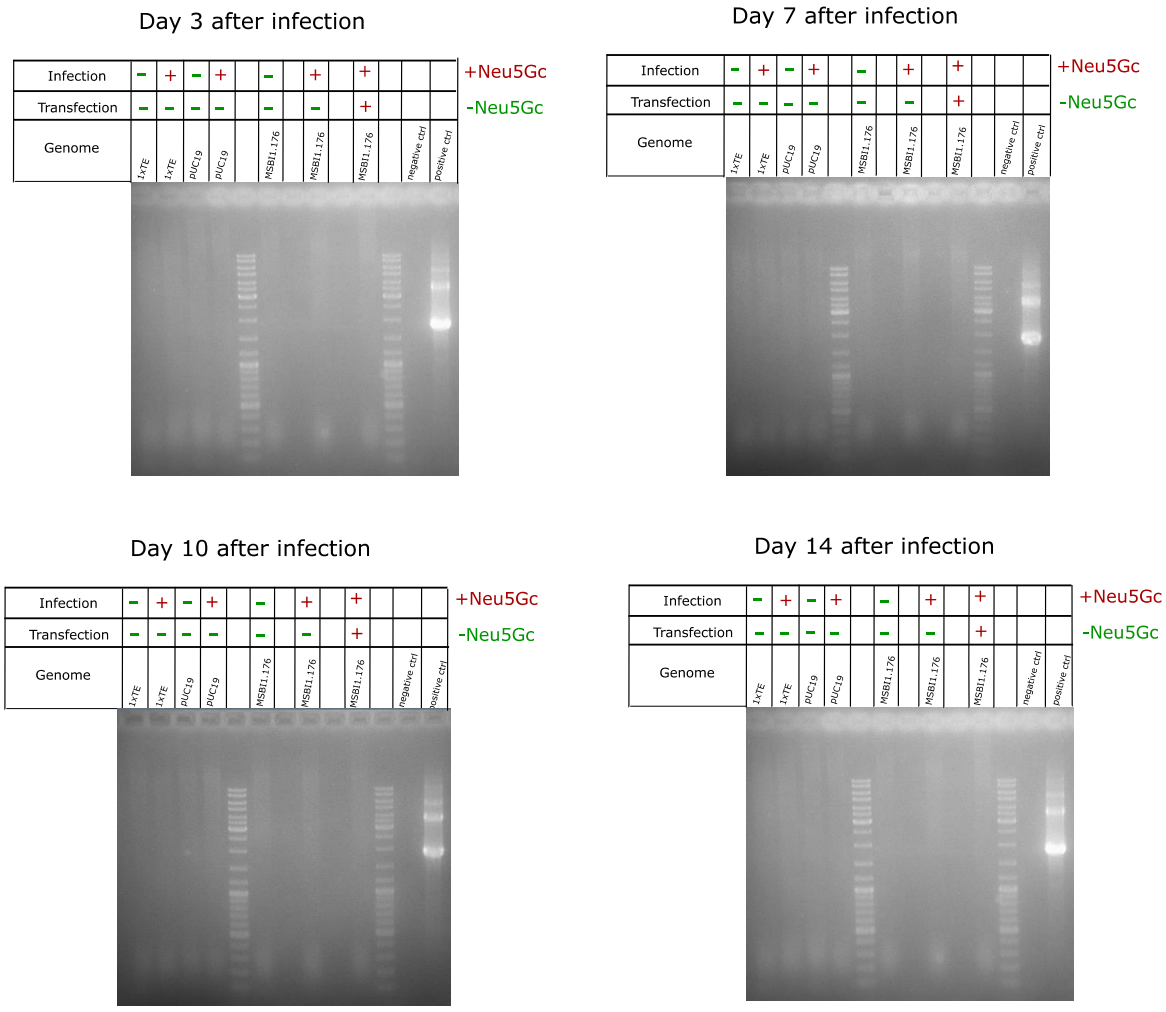


Figure 3.17 PCR on DNA extracted from L-1236 cells after infection with supernatants from transfection with MSBI1.176 and controls using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction. The red color letters indicate the

conditions with 100-fold excess of Neu5Gc and the green color the ones without addition of external Neu5Gc.

3.3.5 Cell viability and cell proliferation assay

In order to assess BMMF effects on cell viability and proliferation, cell viability and proliferation assays were performed after transfection of L-428 cells with linear MSBI1.176, circular MSBI1.176, an overexpression (OE) plasmid encoding for MSBI1.176 Rep wild type (WT) protein or a MSBI1.176 Rep aggregation mutant protein (MUT).

Cell viability was measured with the CellTiter-Glo® Luminescent Cell Viability Assay kit (Promega) (**Figure 3.18)**. However, no significant impact on cell viability was observed when comparing each condition with the corresponding control reactions.

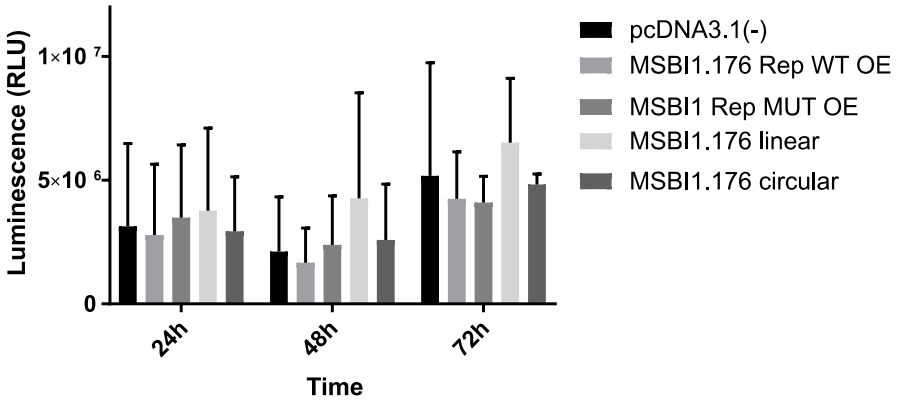


Figure 3.18 Luminescent signal after transfection of L-428 cells with pcDNA3.1(-) control vector, MSBI1.176 Rep WT OE, MSBI1.176 Rep MUT OE, MSBI1.176 linear and MSBI1.176 circular genome measured at three different time points post transfection.

Cell proliferation was tested using the Click-iT™ Edu Plus Alexa Fluor™ 647 Imaging Kit (Thermo Scientific) (**Figure 3.19)**. No significant

difference in cell proliferation was observed when comparing cells overexpressing either wild type MSBI1.176 Rep protein or mutant MSBI1.176 Rep protein with the corresponding controls at the different time points.

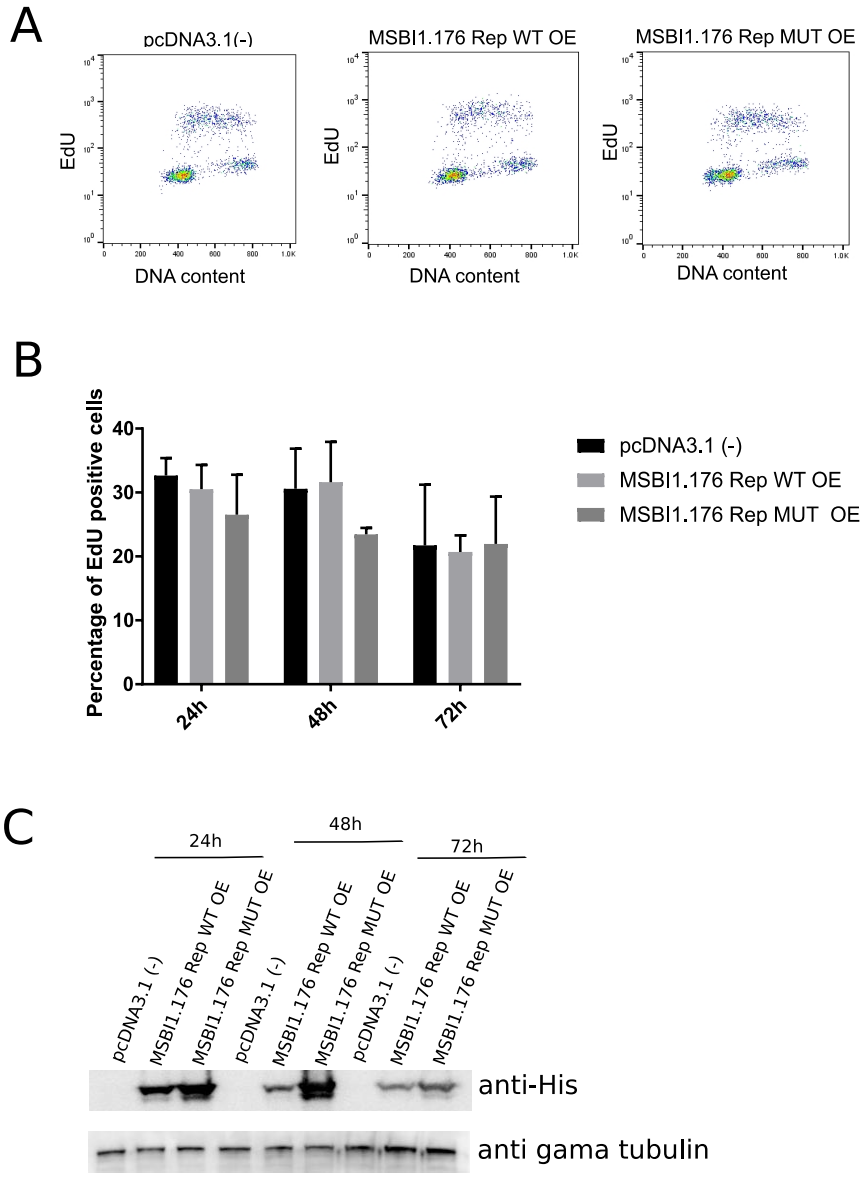


Figure 3.19 Summary of cell proliferation assay in L-428 cells after transfection with pcDNA3.1(-) control vector, MSBI1.176 Rep WT OE protein and MSBI1.176 Rep Mut protein. **A)** representative EdU positive cells at 48h post transfection, **B)** percentage of EdU positive cells after 24h, 48h and 72h for all three conditions, **C)** western blot analysis suing anti-His Rep antibody and alpha tubulin as loading control. No difference was observed for the two genomes.

3.4 Serological tests

A link of BMMFs to breast cancer or to Hodgkin's disease etiology could lead to a detectable immune response to those agents specifically in patients suffering from the respective disease. In this case elevated antibody titers directed against BMMF gene products in sera of disease patients would be expected compared to healthy controls.

To test the presence of serum antibodies (IgG in this study) in breast cancer patients, sera from breast cancer as well as Hodgkin's Lymphoma patients were analyzed by ELISA assays in order to identify a potential immune reaction of those patients against the Rep protein of the MSBI1.176 isolate (which is also highly conserved in the majority of the remaining identified agents of BMMF group1). The wild type MSBI1.176 rep protein was used as antigen and EBNA1 protein of Epstein Barr Virus (EBV) as a positive control and co-marker. A specificity of anti-Rep reactivity in human plasma samples has already been previously reported (Eilebrecht, Hotz-Wagenblatt et al. 2018).

Out of 50 samples from patients suffering from Hodgkin's lymphoma, no significant difference in seropositivity was observed as compared to plasma samples from healthy controls. Interestingly, for the 30 samples from breast cancer patients that have been analyzed significant difference was detected when compared to healthy plasma samples (**Figure 3.20**). The same samples were tested for EBNA1 response, but no significant difference was observed between the seropositivity of the disease samples in comparison to the plasma controls (**Figure 3.21**).

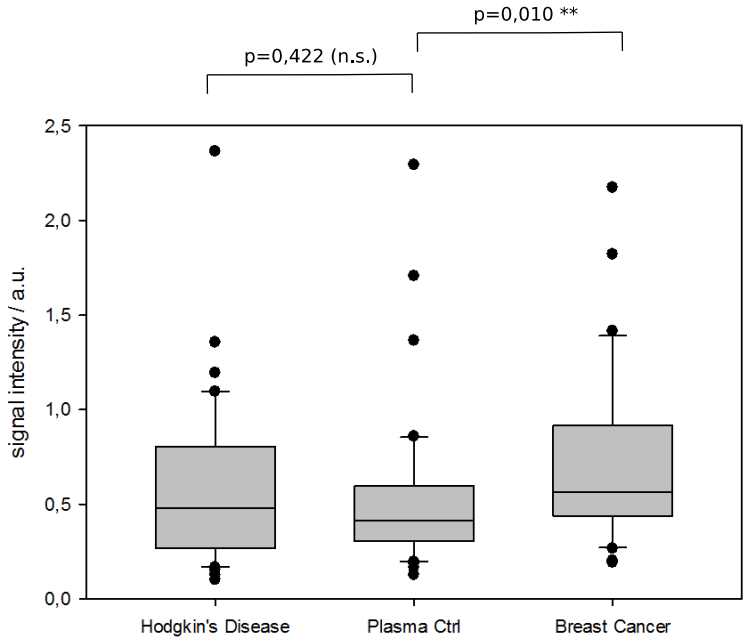


Figure 3.20 Box plot analysis based on ELISA results. Comparison between 50 Hodgkin's Lymphoma serum samples, 30 breast cancer serum samples and 30 plasma controls using Wild type (WT) Rep MSBI1.176 protein as antigen.

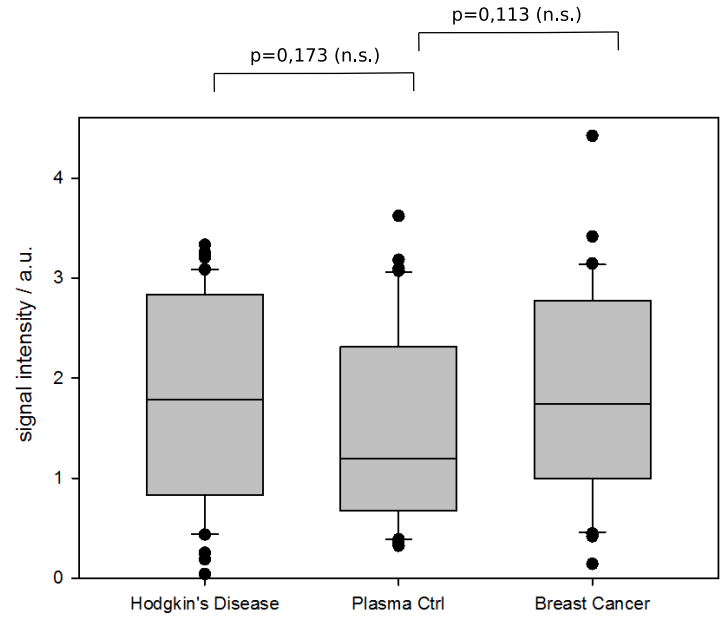


Figure 3.21 Box plot analysis based on ELISA results. Comparison between 50 Hodgkin's Lymphoma serum samples, 30 breast cancer serum samples and 30 plasma controls using the EBNA1 protein as antigen.

Additionally, plasma samples from female donors were compared with plasma samples from male ones, testing for gender- and age-based differences in antibody response. The corresponding data reveal no significant difference between the two genders (**Figure 3.22**). Also for the age, no significant correlation was observed with the strength of the antibody response.

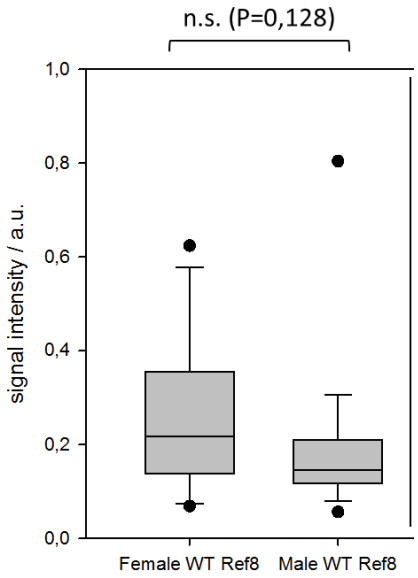


Figure 3.22 Box plot analysis based on ELISA results. Comparison between plasma controls from female and male donors using the Wild type (WT) Rep MSBI1.176 protein as antigen.

4. Discussion

Breast cancer is the most common cause of cancer mortality in women worldwide (Akram, Iqbal et al. 2017). Several factors are correlated with an increased risk of the disease (so called “risk factors”) including sex, age, family history and genetics, reproductive and menstrual history, endogenous and exogenous hormones, as well as personal and lifestyle factors, such as body size and weight, diet and nutrition (National Breast and Ovarian Cancer Center 2009). According to diet as risk factor, more specifically, a number of epidemiological studies indicate a potential relation between the consumption of red meat and milk and an increased incidence of breast cancer (Boyd, Stone et al. 2003, Moorman and Terry 2004, Taylor, Burley et al. 2007, Linos, Willett et al. 2008, Farlow, Xu et al. 2009, Kabat, Cross et al. 2009, Maruyama, Oshima et al. 2010, Farvid, Cho et al. 2015). In addition, it has been previously reported that species-specific risk factors might contribute to the development of breast cancer as well as colon cancer and neurodegenerative diseases like Parkinson’s disease and Multiple Sclerosis (zur Hausen and de Villiers 2015, zur Hausen, Bund et al. 2017). Looking closely into the dietary habits globally, as analyzed in these reviews, it is intriguing that dairy cattle is most prevailing in almost all high risk regions for development of breast and colon cancer, such as North America, most of the European countries and Australia, while in low risk countries, such as Mongolia and Bolivia, different cattle species are found (e.g. Yak or Zebu) (zur Hausen and de Villiers 2015). On the other hand, several studies suggest that viruses might play a role in the development of breast cancer, amongst them Human Papilloma Viruses (HPV), Epstein-Barr Virus (EBV), Mouse Mammary Tumor Virus (MMTV) and Bovine Leukemia Virus (BLV) (Akhter, Ali Aziz et al. 2014, Lawson, Glenn et al. 2016, Nartey, Mazzanti et al. 2017, Gannon, Antonsson et al. 2018, Lawson, Salmons et al. 2018), however the experimental and clinical data towards a causal association of viruses and cancer are controversial as it is highlighted by different studies (Joshi and Buehring 2012, De Paoli and Carbone 2013). The isolation of circular

episomal DNA agents from commercially available milk samples, healthy cattle blood samples and human blood and tissue led in the postulation of “bovine milk and meat factors” (BMMFs) being transmissible to humans and potentially playing a role in the pathogenesis of breast and colon cancer (zur Hausen, Bund et al. 2017). The authors of this review suggest that epidemiological data link several types of cancers, among them Hodgkin’s disease, as well as neurodegenerative, cardiovascular and autoimmune diseases with the consumption of red meat and dairy products.

It is assumed that an infection with these agents takes place early in life, probably during the weaning period and by the exposure of children to the commercially available cow milk (zur Hausen, Bund et al. 2017). This theory is supported by studies indicating a highly protective role of human milk, which contains specific sugars apparently blocking receptors that act as binding sites of several agents or generally lower amounts of risk factors which are only present in cow products. These agents may persist latently for several decades and cause chronic inflammatory lesions, resulting in mutagenic oxygen and nitrogen radicals. BMMF activation later in life could be due to various yet unknown synergistic interactions, such as infection by different viruses from the Herpes virus family.

A number of questions are raised based on these considerations:

1. How can the existing epidemiological data be linked with the molecular biology of breast cancer?
2. Can these agents of obviously bacterial origin replicate in human cells?
3. Are these agents capable of inducing a human immune response?

Within this thesis a previously unknown BMMF sequence was identified by Next Generation Sequencing (NGS) from silico breast tissue samples. Interestingly, it was shown that bacterially derived BMMFs can actively persist and replicate in human pre-B cells and that an antibody response against the

replication protein of BMMFs can be detected in serum samples of breast cancer patients.

- **Identification of three new BMMFs**

Based on epidemiological data linking the consumption of red meat and dairy products and the incidence of breast and colon cancer and neurodegenerative diseases, commercially available cow milk, bovine serum samples and blood and tissue samples from Multiple Sclerosis patients were analyzed resulting in the isolation of 18 circular DNA agents, mainly from bovine samples (Funk, Gunst et al. 2014, Gunst, Zur Hausen et al. 2014, Lamberto, Gunst et al. 2014, Whitley, Gunst et al. 2014). These molecules have been organized into four groups based on their nucleotide similarity to previously identified circular genomes. All contain a large open reading frame encoding for a putative replication initiator protein (rep protein). In addition, they reveal an A/T-rich region and four direct repeat sequences up-stream of the Rep ORF, characteristic for circular Replication-associated protein (Rep)-encoded single-stranded (CRESS) DNAs (Rosario, Duffy et al. 2012). CRESS DNA viruses display considerable sequence diversity and their potential pathogenicity in certain hosts and the environment is poorly understood (Fernandez, Nunez-Ramirez et al. 2016).

An important part of this PhD thesis was the continuing analysis of a commercially available milk samples as well as its expansion to different dairy products. This approach resulted in the identification of two previously unknown DNA agents from milk that showed similarity to one of the previously identified agents. In addition, three of the previously isolated agents, were reproductively identified in different dairy samples. The use of additional primers, which are specific for the BMMF group 2 isolated, has also revealed approximately 30 new molecules coming from different dairy products (card cheese, yogurt, crème fraiche) and milk samples indicating that there is significant diversity of such agents. Repetitive re-isolations of BMMF group 1

representatives might indicate less diversity for this group in milk when compared to BMMF2 isolations. BMMF isolation from milk and yogurt, representing dairy products with a longer processing phase, in general, might hint at significant stabilization of BMMFs surviving even longer production cycles and longer self-life observed for these products. Based on recently published data (Eilebrecht, Hotz-Wagenblatt et al. 2018), a co-infection with more than one of these agents might be responsible for the pathogenicity of BMMFs, as replication of the smaller genome MSBI1.176 was increased when co-transfected with the larger genome CMI1.252 in HEK293TT cells in comparison to the single-transfection condition of each isolate alone. In any case, these agents might represent novel, still not fully characterized pathogens and therefore their specific pathogenicity model deserves analysis from various aspects.

Assuming a role of BMMF infection in direct carcinogenesis during breast cancer formation, the corresponding BMMF genomes might be identifiable by sequencing from total DNA isolated from breast cancer. It has been postulated that BMMFs are transferred to their target cells after initial infection early in life (zur Hausen, Bund et al. 2017). BMMFs were discovered by polymerase chain reaction (PCR) using specific primers based on previously characterized agents. In this thesis, a different approach was developed in order to *de novo* identify episomal DNA molecules from any type of sample without any prior sequence knowledge. Therefore, Next Generation Sequencing (NGS) was used to analyze breast cancer biopsies. Metagenomic approaches can detect known or unknown viruses and they do not require prior sequence information. This approach is highly sensitive and it can quickly recover full viral genome sequences with less initial material in comparison to the classical methods of cloning. When combined with nucleic acid sequence-independent amplification techniques (such as Rolling Circle Amplification, RCA), NGS can enhance sensitivity.

The step of sample preparation is very crucial for downstream NGS approaches. Eukaryotic nucleic acids can obscure the isolation and identification of viral and bacterial DNA and therefore it is beneficial to remove

this DNA prior to sequencing. Additionally, removal of host genetic material is also important for increasing the detection sensitivity of the NGS platform. Several virus enrichment approaches have been developed over the years, including sequence independent single-primer amplification (SISPA), virus discovery cDNA-amplified fragment length polymorphism (VIDISCA) and RCA. The selection of an appropriate high-throughput sequencing technique is also essential when performing such experiments. The longer the obtained sequence reads, the better they can be used for *de novo* assembly and generation of contigs, enhancing the possibilities of succeeding in nucleotide databank searches. In addition, the analysis of the excessive amount of data obtained from NGS is the most demanding step of metagenomic analyses.

Taken these aspects into account, prior to the analysis of the breast cancer biopsies, a protocol was established for the reliable purification of plasmid DNAs from tissue samples and their subsequent detection by NGS. Two kits, one for isolation of total DNA and one for isolation of plasmid DNA, were used for the DNA isolation from breast cancer tissue biopsies, followed by Rolling Circle Amplification (RCA) for amplification of small single-stranded DNAs and Plasmid-Safe DNase digestion for removal of chromosomal DNA. Eleven breast tissue samples and their corresponding tissue adjacent to tumor (TAT) samples were used for DNA isolation with both kits. However, due to the large amount of adipose tissue in the TAT samples and low amounts of DNA obtained from them with the plasmid kit, further steps of quality control and library preparation for NGS were performed only on the total DNA obtained with the other kit. A protocol for FFPE samples has been used for the library generation of the fourteen samples (seven breast cancer tumor samples and seven corresponding TAT samples) and sequencing was performed in a Illumina HiSeq4000 platform initiating paired-end reads of 100 bp length.

The bioinformatical processing was performed according to a recent study (Jorgensen, Xu et al. 2014). Briefly, after removal of adapter sequences and low quality nucleotides, the Illumina reads were mapped to the human genome 38. The unmapped reads were assembled *de novo* using IDBA-UD

and parameters according to the above mentioned study. Therefore implicitly all putative bacterial sequences which are normally filtered out by standard algorithms were still present in the reads. At the next step, a pipeline was developed to search for circular contigs, since it was expected that if a circular contig is sequenced with high enough coverage and assembled into a single contig, reads overlapping the ends of the contig should exist. The potentially circular contigs were mapped again to the human genome and BLAST search against the previously identified genomes (Funk, Gunst et al. 2014, Gunst, Zur Hausen et al. 2014, Lamberto, Gunst et al. 2014, Whitley, Gunst et al. 2014) was performed. One circular contig of 1620 bp size was found in a TAT sample, which showed 85% similarity to the previously isolated MSSI.162 (Gunst, Zur Hausen et al. 2014) according to BLAST search. The significant differences between the two genomes exclude the possibility of a laboratory contamination. MSSI1.162 belongs to the Group 4 of our isolates and it is closely related to a plasmid of the *Psychrobacter*. Previously tested, it showed substantially less transcriptional activity compared to the other isolates (unpublished data). This contig reveals also 95% similarity to an uncultured bacterial plasmid isolated from plasma of patients in maintenance hemodialysis (Biagini, P), raising the question whether such agents may play a role in the pathogenesis of other diseases as well. Interestingly, it was present in one of the TAT samples, which could indicate that BMMFs as causative agents are more prevalent in TAT than in tumor, possible due to lost after tumorigenesis, because no DNA replication is taking place anymore. Even though the breast tissue samples that have been used in this study had been stored for many years and subjected to several cycles of freezing and thawing, the success of one out of the fourteen samples analyzed suggests that the developed approach is adequate for the discovery of such sequences and that the use of fresh tissue samples could reveal more positive results in future attempts.

- **Replication of BMMFs**

Breast cancer is not the only disease whose incidence correlates with the consumption of meat and dairy products. Several epidemiological data indicate a potential correlation between milk and read meat consumption as well as the animal exposure and the development of Hodgkin's disease and other cancers and neurodegenerative diseases (zur Hausen, Bund et al. 2017). In addition, 24 full-length Torque Teno virus types (a single-stranded DNA virus, with similar structure however different nucleic acid composition) have been isolated from a single spleen biopsy sample from a patient with Hodgkin's lymphoma (Jelcic, Hotz-Wagenblatt et al. 2004).

To specifically screen for a potential permissive cell culture system for BMMFs, two Hodgkin's Lymphoma cell lines, the L-1236 and L-428 cell lines, were selected for transfection with the Multiple Sclerosis brain isolate MSBI1.176, which after transfection revealed the strongest transcriptional activity and protein expression (Eilebrecht, Hotz-Wagenblatt et al. 2018). These cells are considered pre-B cells (Kuppers, Rajewsky et al. 1994). Blood cells of the B cell lineage may transport BMMFs to the affected organ. Previous experiments (zur Hausen, unpublished data) showed small putative particles in the L-1236 cells after infection with herpes simplex virus (HSV-1), which may hint at specific viral infections playing a role in the pathogenesis of Hodgkin's disease (reactivation by HSV-1 infection).

After transfection of both cell lines with the MSBI1.176 genome, DNA was harvested at different time points. Digestion with DpnI restriction digestion was included before further analysis, in order to remove methylated input DNA while retaining non-methylated replicated DNA. However, several points should be taken into account for a reasonable assessment of replication competence: i) DpnI restriction digestion efficacy is never at 100%, ii) only a limited number of DpnI sites is covered by primers in qPCR analyses, rendering the detection of newly synthesized DNA less sensitive, and iii) DNA methyltransferases, a family of enzymes that catalyze the methylation of DNA and shifts methyl groups from one nucleotide to another,

can bias the restriction efficiency of the DpnI sites, interfering with the detection of replicated DNA. Analyses by long-PCR and qPCR showed persistence of the MSBI1.176 genome up to day 36 post transfection. The presence of newly synthesized (non-methylated) DNA pointed at replication of the transfected genome. Several factors may play a role in MSBI1.176 replication, such as a specific activity of the Rep protein or of host cellular replicases, as well as the AT-rich region, which is a major region for controlling the DNA replication initiation. Further experiments with specific mutations and testing for responsible elements would provide more information.

It has previously been reported that in co-transfection experiments of HEK293TT cells with MSBI1.176 and CMI1.252 genomes (Eilebrecht, Hotz-Wagenblatt et al. 2018), CMI1.252 complements the replication of MSBI1.176. However, this effect was not observed during co-transfection of the L-428 cells with these genomes potentially due to the high rate of MSBI1.167 replication already in the single transfection or due to differences in transcriptional activity between the different cell lines. Translational differences or transfection efficiency could also influence the outcome. CMI1.252 showed stable replication up to 28 days post transfection in this cell line as well, similar to the replication of TTVs in 293TT cells (Borkosky, Whitley et al. 2012) and similar to CMI1.252 replication in HEK293TT cells (Eilebrecht, Hotz-Wagenblatt et al. 2018).

Very interestingly, within routine replication experiments additional repeatedly-observed bands were identified. Further co-transfection experiments involved the transfection with small genomes of 613 bp length, isolated after transfection of L-428 cells with MSBI1.176. Long-PCR on DpnI digested DNA revealed these small extra bands representing the MSBI1.176 genome with substantial deletions (at 1700 bp for MSBI1.176). These molecules may also arise from defective replication of the initial genome. In order to test for potential functional importance of these molecules, replication and stability were investigated and they were tested for regulatory effects. These molecules might have an interfering or helper effect during replication

of the MSBI1.176 genome from which they originated. Such an effect is established for Begomoviruses, a genus of the family of Geminiviruses (plant animals with circular, single-stranded DNA genome of ~ 2.7 kb) that their genome is accompanied by small circular satellite DNAs, which can replicate autonomously or transreplicate by their non-cognate helper virus (Rosario, Duffy et al. 2012, Hanley-Bowdoin, Bejarano et al. 2013, Rizvi, Choudhury et al. 2015). In the case of MSBI1.176 sg613, the structure and autonomous replication, as well as any helper or regulatory effect for DNA replication of the small isolated genome were analyzed. This genome consists of the last part of the repeat region up stream of the Rep ORF and continues covering approximately 2/3 of the latter, remained stable in this rearranged form (de Villiers, unpublished). Analysis of the DpnI-persistent DNA by long PCR and qPCR revealed that this small genome is capable of independent replication, suggesting that only the last iteron of the repeat region or control elements in the residual sequence might be sufficient to initiate replication. It remains elusive whether the C-terminus of the Rep protein, which is lacking in the sg613, plays a role in aggregation formation and replication. Interestingly, Torque-Teno viruses (TTVs) of approximately 1000 bp length have also shown to persist in different cell lines for more than 28 days post transfection (de Villiers, unpublished data). However, in the L-428 cell line, no clear effect was observed after co-transfection with the small genome and MSBI1.176, indicating that the minimized genome is competent of autonomous replication, but any competition role with the full length MSBI1.176 genome requires further investigation. It would also be important to test whether different parts of the iterons might affect replication of MSBI1.176 in an abortive or productive way, possibly by an induction of a Rep dimerization-monomerization process, since it has been demonstrated that the binding of short dsDNA oligonucleotides including the iterons affects the *winged-helix* WH1 domain of the *Pseudomonas* pPS10 plasmid promoting the assembly of different structures (such as aggregates and amyloid spheroids) instead of the usual monomerization that takes place and leading to an inhibition of plasmid replication (Fernandez, Nunez-Ramirez et al. 2016).

- **Study of Neu5Gc glycoproteins and phenotypic characterization**

As replication of BMMFs in pre-B cells was proven, the next question dealt with potential cellular receptors mediating a BMMF infection. It has already been reported that the incorporation of N-glycolylneuraminic acid (Neu5Gc) into human cell glycoprotein receptors leads to chronic inflammatory reactions. This has led to a model, in with such a Neu5Gc incorporation potentially promotes cancer and other diseases (Alisson-Silva, Kawanishi et al. 2016). Neu5Gc cannot be synthesized in humans due to an exon deletion in the enzyme responsible for the formation of Neu5Gc from its precursor sialic acid N-acetylneuraminic acid (Neu5Ac). Neu5Gc is metabolically incorporated via dietary consumption specifically of red meat and cow milk. A recent publication (zur Hausen, Bund et al. 2017) summarizes a number of reasons supporting the theory that a functional receptor for BMMFs is established by Neu5Gc incorporation. These include reports that Neu5Gc acts as a specific receptor for several viruses, like human polyoma virus 9, canine and fenine parvoviruses, types of influenza A and parainfluenza viruses. In order to test an involvement of Neu5Gc in increasing the permissiveness of human cells for BMMF replication, we compared a condition of MSBI1.176 transfection with external addition of 100 µM free Neu5Gc (approximately 100-fold excess compared to the standard Neu5Gc in 10% FCS cell culture medium) with one without addition. However, no significant difference in MSBI1.176 genome replication was detected.

Furthermore, infection experiments of fresh cells with MSBI1.176-positive supernatants of MSBI1.176-transfected cells were performed from the first time in a further attempt to test an involvement of Neu5Gc and to establish a general infection system for BMMFs, but no infectivity was observed. These negative results might be due to the experimental setting used, since it disregarded the presence of Neu5Gc in fetal calf serum added to the cell culture, which might repress the binding of the particles to the receptors by binding to the particle surface. In addition, with this setting there is no direct contact of the target and producer cells and there is a long time of

separation of production of viral particles and infection of fresh cells, while at the same time centrifugation and filtering steps cause general instability and loss of viral particles. Furthermore, the freezing and thawing steps of supernatants might have already affected the stability of protein particles and vesicles, which might decrease infectivity. Therefore, infectivity of BMMFs still remains an open question and a more appropriate experimental set-up could provide conclusive answers for a potential role of Neu5Gc-modified receptors for infectivity of BMMFs. In addition, focusing on different combinations of agents and not on single MSBI1.176 infections might provide useful information and broaden our knowledge about the role of these agents in human diseases.

Expression/replication of putative pathogens phenotypic characterization is important. It has already been reported that several cancer-related genes are significantly regulated by MSBI1.176 and CMI1.252, suggesting a potential link of BMMFs to cellular pathways damaged during carcinogenesis (Eilebrecht, Hotz-Wagenblatt et al. 2018). Cell viability and cell proliferation assays allow the monitoring of cell growth progression in vitro and provide multiplexed information about cell cycle duration, transition points, proliferative quiescence, etc. (Simone, T. et al. 2009). The main focus was to test for differences in cell viability and recovery as well as to determine cell health upon BMMF genome transfection. Over-expressed proteins, wild type and aggregation mutant of MSBI1.176 Rep protein, were used as controls. No phenotypic changes were observed with either of the approaches. Besides, when these agents act as indirect carcinogens, no phenotypic changes should emerge. The absence of phenotypic changes indicates that there is no difference between the two over-expressed genomes upon transfection and that the function/activity of Rep protein expression does not play a pivotal role in manifestation of phenotypic changes of viability and cell progression. This result could also be an effect of the promoter used (in this case cytomegalovirus, CMV) which is a strong one and potentially affects cell proliferation. The use of a different promoter is recommended for future experiments, such as the human Ubiquitin C

promoter (UBC) or the mouse phosphoglycerate kinase 1 promoter (PGK), as well as the use of different DNA amounts used for transfection.

- **MSBI1.176 Rep antibody response in human serum samples**

Transcription, translation and replication are important aspects of bioactive pathogens. The expression of the corresponding antigens in the host will then raise an immune response consisting of a production of antigen-specific antibodies. Thus, bioactivity of pathogens might also be deduced from the detection of specific host antibodies targeting the putative infectious agents, indicating that human individuals have been exposed to such agents.

An involvement of BMMFs in breast cancer or Hodgkin's disease etiology should lead to a detectable immune response to those agents specifically in patients suffering from the respective disease. In this case elevated antibody titers directed against BMMF gene products in sera of disease patients would be expected compared to healthy controls. Therefore, sera from patients with breast cancer and Hodgkin's Lymphoma were analyzed by ELISA assays in order to identify a potential immune response of these patients against the Rep protein of the MSBI1.176 isolate.

Although a detectable immune response against the MSBI1.176 Rep antigen was observed in Hodgkin's disease samples, the signal intensity was low compared to other cohorts and no statistically significant difference was observed between the patients and the healthy controls This decreased antibody levels might hint at no immunogenicity of BMMFs or at possible partial immunotolerance due to a very early acquisition of these types of infections. On the other hand, breast cancer serum samples revealed a significant difference (P = 0,010) when compared to healthy plasma controls, supporting the theory of an infection with BMMF-like agents early in life and neutralizing antibody formation. However, these results have to be

reproduced comparing breast cancer serum samples with serum samples of matched healthy controls, in order to rule out that the observed differences in immune response result from the comparison of plasma with serum.

Conclusions

Epidemiological data indicate a potential relation between the consumption of bovine meat and dairy products and the incidence of breast cancer and colon cancer, as well as neurodegenerative diseases. A model in which species-specific risk factors might contribute to the etiology of such diseases has been reported (zur Hausen and de Villiers 2015). Based on this model, circular DNA molecules have been isolated from commercially available milk and dairy products during this study. In addition, the isolation of a similar molecule, following a High Throughput Sequencing protocol that has been developed as part of this thesis, from a tissue sample adjacent to breast cancer tissue opens new starting points for investigation and analysis.

This study established a long-term persistence and replication of a set of bacterial-derived plasmid-like agents in two Hodgkin's disease-derived cell lines. Co-transfection experiments with more than one agent failed to prove regulatory effects in these cell lines likely to already individual replication levels, even though helper effects for two of them has been shown previously by our group in a different cell line (Eilebrecht, Hotz-Wagenblatt et al. 2018).

The demonstration that the incorporation of Neu5Gc into human cell membranes can cause an immune reaction leading to chronic inflammation steered the model of a potential role of Neu5Gc as receptor for BMMF infection, especially since it has already been reported that it acts as receptor for several viruses. The experiments performed in this study only showed a slight increase of replication after external addition of Neu5Gc, but they did not serve conclusive results regarding infectivity of BMMFs and additional investigations under different conditions are required.

Finally, the elevated anti-MSBI1.176 Rep immune response detected in sera from breast cancer patients compared to healthy donors indicated that infection with BMMF related agents resulted in neutralizing antibody formation.

This PhD thesis focuses on an epidemiologically suggested involvement of dietary habits as a risk factor for the formation of breast cancer and provides important information on a new and yet unknown field. However, there are still open questions that will require additional studies.

5. Summary

Breast cancer is the second-leading cause of cancer-related deaths in women worldwide. A number of factors are associated with an increased risk of the disease. Epidemiological data indicate a potential link between the consumption of bovine meat and dairy products and the incidence of breast cancer, as well as other types of cancer such as colon and prostate cancer and neurodegenerative or autoimmune diseases. Analysis of the epidemiological data led to the suggestion of a new theory introducing a putative role of species-specific factors originating from the species *Bos taurus* contribute to the etiology of these diseases. Based on these speculations, our group started analyzing bovine serum samples, commercially available milk, as well as human blood and tissue samples from Multiple Sclerosis patients. This study resulted in the isolation of 18 circular DNA agents, so called Bovine Meat and Milk Factors (BMMFs), which have been divided in four groups according to their similarity to known DNA molecules. This PhD thesis focused on continuing this study, by expanding it in different dairy products such as yogurt, crème fraiche, butter and card cheese, resulting in the isolation of two novel DNA molecules which share similarity to some extend with the MSBI1.176 isolate.

In addition to the classical molecular biology-based DNA isolation techniques, also isolation of DNA from breast cancer tissue samples by Next Generation Sequencing (NGS) was performed. A protocol was established for the reliable purification of plasmid DNAs from mammalian cells. DNA from seven breast cancer tissue samples and their corresponding Tissue Adjacent to Tumor (TAT) samples was analyzed by NGS resulting in the isolation of a circular DNA molecule from a TAT sample, which showed 85% nucleotide identity to the previously isolated MSSI1.162 and 95% similarity to an uncultured bacterium plasmid isolated from plasma of patients in maintenance hemodialysis. This discovery sets new starting points for further investigation and analysis.

Epidemiological data suggest a potential link between the consumption of red meat and animal exposure and the incidence of Hodgkin's disease, while in the past 24 full-length Torque Teno viruses (a single stranded DNA virus) have been isolated from the spleen biopsy of a single patient with Hodgkin's lymphoma. Therefore, searching for a permissive cell system for the identified BMMFs, two Hodgkin's disease cell lines, the L-1236 and L-428, were selected for transfection and long-term persistence and replication of a set of bacterially-derived plasmid-like agents was established. Co-transfection experiments with more than one agent failed to prove complementation effects in these cell lines, even though helper effects for two of them have been shown previously by our group in the HEK293TT cell line.

The demonstration that the incorporation of N-Glycolylneuraminic acid (Neu5Gc) into human cell membranes can cause an immune reaction leading to chronic inflammation steered the model of a potential role of Neu5Gc as receptor for BMMF infection, especially since it has already been reported that it acts as receptor for several viruses. The experiments performed in this study only showed a slight increase of replication after external addition of Neu5Gc, but they did not serve conclusive results regarding infectivity of BMMFs and additional investigations under different conditions are required.

Finally, sera from breast cancer and Hodgkin's disease patients were analyzed by ELISA assays in order to identify a potential immune reaction of those patients against the Rep protein of MSBI1.176. The elevated anti-MSBI1.176 Rep immune response detected in sera from breast cancer patients compared to plasma from healthy donors indicated that infection with BMMF related agents resulted in measurable antibody formation. However, no difference in seropositivity was observed when serum samples from patients suffering from Hodgkin's lymphoma where compared to healthy plasma samples.

This PhD thesis focuses on an epidemiologically suggested involvement of dietary habits as a risk factor for the formation of breast cancer and provides important information on a new and yet unknown field. However, there are still open questions that require further studies.

Zusammenfassung

Brustkrebs ist die zweithäufigste Ursache Krebs-assoziierter Todesfälle bei Frauen weltweit. Viele verschiedene Faktoren sind mit diesem Krankheitsrisiko assoziiert. Epidemiologische Daten deuten auf eine potenzielle Verbindung zwischen dem Verzehr von Rindfleisch als auch Milchprodukten und der Brustkrebsinzidenz hin, wobei auch andere Arten von Krankheiten wie sowohl Dickdarm- und Prostatakrebs als auch neurodegenerative oder Autoimmundefekte damit in Verbindung gebracht werden. Aufgrund der epidemiologischen Daten wurde eine neue Hypothese aufgestellt, die Rinderarten, die sich aus der Gattung *Bos taurus* ableiten, einen spezies-spezifischen Faktor zuschreibt. Dieser Faktor sollte für die Ätiologie der genannten Krankheiten (mit)verantwortlich sein. Darauf aufbauend analysierte unsere Arbeitsgruppe sowohl Rinderseren und kommerzielle Milchproben wie auch humanes Blut und Gewebe von Patienten mit Multipler Sklerose. In dieser Studie wurden 18 zirkuläre DNA-Moleküle isoliert, genannt bovine milk and meat factor (BMMF), die aufgrund ihrer Sequenzähnlichkeit zu bereits bekannten DNA Molekülen in vier Untergruppen klassifiziert wurden. Diese Dissertation führt DNA-Isolationen fort, indem weitere Milchprodukte wie Joghurt, Creme Fraiche, Butter und Quark untersucht wurden. Hieraus konnten zwei neue DNA-Moleküle isoliert werden, die eine Ähnlichkeit zu einem bereits aus früheren Studien bekannten Molekül zeigen (MSBI1.176).

Zusätzlich zur klassischen molekularbiologischen Isolation und Analyse von DNA wurden hier Brustkrebsgewebe mittels „Next-Generation-Sequenzierung“ (NGS) auf das Vorhandensein von BMMFs getestet. Dafür wurde ein reproduzierbares Protokoll zur Aufreinigung von Plasmid-DNA aus Säugetierzellen etabliert. Die so gewonnene DNA aus sieben Brustkrebsgeweben und gesundem Tumor-benachbarten Kontrollgewebe wurde mittels NGS analysiert. Aus diesem Kontrollgewebe wurde ein zirkuläres DNA Molekül isoliert, welches 85% Sequenzidentität auf Nukleinsäureebene zu einem bereits aus erster Studie bekanntem Isolat MSSI1.162, zeigt und auch 95% Ähnlichkeit zu einem bakteriellen Plasmid

eines Hämodialyse-behandelten Patienten aufweist. Dies zeigt neue und interessante Ansätze für weitere Untersuchungen auf.

Epidemiologische Daten legen auch eine potenzielle Verbindung zwischen dem Verzehr und der verstärkten Exposition mit rotem Fleisch und Hodgkin Lymphomen nahe. Darüber hinaus wurden in der Vergangenheit 24 Torque-Tenoviren (einzelsträngige DNA Viren) aus Milzbiopsien eines Hodgkin Lymphom-Patienten isoliert. Daher wurden für die Suche nach einem permissiven Zellkultursystem für die identifizierten BMMFs zwei Zelllinien, die ursprünglich aus Hodgkin-Lymphomen stammen, getestet (L-1236 und L-428). Nach transienter Transfektion dieser Zelllinien konnte eine Langzeit-Persistenz und Replikation der verschiedenen Plasmid-ähnlichen Agenzien in diesen beiden Zelllinien aufgezeigt werden. In Ko-Transfektionsexperimenten mit Kombinationen zweier unterschiedlicher BMMFs wurde jedoch keine Komplementation der Agenzien in Hodgkin Zelllinien beobachtet, obwohl solche Helfereffekte in der Nierenzelllinie HEK293TT in unserer Arbeitsgruppe gezeigt wurden.

Der Einbau des Zuckermoleküls N-Glycolylneuraminsäure (Neu5Gc) in humane Zellmembranen kann Immunreaktionen und dadurch chronische Entzündungen verursachen. Darüber hinaus dient es als Rezeptor für verschieden Viren, weshalb Neu5GC eine potenzielle Rolle im Zelleintritt der BMMFs spielen könnte. Die durchgeführten Experimente zeigten allerdings nur eine geringe Verstärkung der Replikation nach externer Zugabe von Neu5GC. Daher lassen sich keine eindeutigen Schlüsse bezüglich der Infektion von BMMFs ziehen und weitere Untersuchungen unter verschiedenen Bedingungen sind hierzu nötig.

Um eine potenzielle Immunreaktion von Patienten mit Brustkrebs und Hodgkin Lymphomen gegenüber dem MSBI1.176-BMMF kodierten Rep-Protein zu untersuchen, wurden ELISA-Experimente durchgeführt. Es wurde eine erhöhte Immunantwort gegenüber MSBI1.176 Rep in Serum von Brustkrebspatienten gegenüber Plasma von gesunden Probanden festgestellt, was eine messbare Exposition gegenüber BMMF zeigt. Allerdings

konnte kein Unterschied der Seropositivität zwischen Serum von Patienten mit Hodgkin-Lymphom und Kontrollplasma beobachtet werden.

Diese Dissertation untersucht eine epidemiologisch basierte Assoziation von diätetischen Gewohnheit als Risikofaktor bei Brustkrebs. Es wurden wichtige Erkenntnisse auf diesem bisher unbekannten Gebiet gewonnen, wobei weitere Studien nötig sind um BMMF eindeutig mit Brustkrebs in Verbindung zu bringen.

6. References

Akhter, J., M. A. Ali Aziz, A. Al Ajlan, A. Tulbah and M. Akhtar (2014). "Breast cancer: is there a viral connection?" Adv Anat Pathol **21**(5): 373-381.

Akram, M., M. Iqbal, M. Daniyal and A. U. Khan (2017). "Awareness and current knowledge of breast cancer." Biol Res **50**(1): 33.

Alisson-Silva, F., K. Kawanishi and A. Varki (2016). "Human risk of diseases associated with red meat intake: Analysis of current theories and proposed role for metabolic incorporation of a non-human sialic acid." Mol Aspects Med **51**: 16-30.

Antoniou, A., P. D. P. Pharoah, S. Narod, H. A. Risch, J. E. Eyfjord, J. L. Hopper, N. Loman, H. Olsson, O. Johannsson, Å. Borg, B. Pasini, P. Radice, S. Manoukian, D. M. Eccles, N. Tang, E. Olah, H. Anton-Culver, E. Warner, J. Lubinski, J. Gronwald, B. Gorski, H. Tulinius, S. Thorlacius, H. Eerola, H. Nevanlinna, K. Syrjäkoski, O. P. Kallioniemi, D. Thompson, C. Evans, J. Peto, F. Lalloo, D. G. Evans and D. F. Easton (2003). "Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies." The American Journal of Human Genetics **72**(5): 1117-1130.

Barzon, L., E. Lavezzo, V. Militello, S. Toppo and G. Palu (2011). "Applications of next-generation sequencing technologies to diagnostic virology." Int J Mol Sci **12**(11): 7861-7884.

Borkosky, S. S., C. Whitley, A. Kopp-Schneider, H. zur Hausen and E. M. de Villiers (2012). "Epstein-Barr virus stimulates torque teno virus replication: a possible relationship to multiple sclerosis." PLoS One **7**(2): e32160.

Bouvard, V., D. Loomis, K. Z. Guyton, Y. Grosse, F. E. Ghissassi, L. Benbrahim-Tallaa, N. Guha, H. Mattock and K. Straif (2015). "Carcinogenicity of consumption of red and processed meat." Lancet Oncol **16**(16): 1599-1600.

Boyd, N. F., J. Stone, K. N. Vogt, B. S. Connelly, L. J. Martin and S. Minkin (2003). "Dietary fat and breast cancer risk revisited: a meta-analysis of the published literature." Br J Cancer **89**(9): 1672-1685.

Buller, R. M., J. E. Janik, E. D. Sebring and J. A. Rose (1981). "Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication." J Virol **40**(1): 241-247.

Chiu, C. Y. (2013). "Viral pathogen discovery." Current Opinion in Microbiology **16**(4): 468-478.

Cliften, P. (2015). Chapter 7 - Base Calling, Read Mapping, and Coverage Analysis A2 - Kulkarni, Shashikant. Clinical Genomics. J. Pfeifer. Boston, Academic Press**:** 91-107.

Coleman, M. P., M. Quaresma, F. Berrino, J.-M. Lutz, R. De Angelis, R. Capocaccia, P. Baili, B. Rachet, G. Gatta, T. Hakulinen, A. Micheli, M. Sant, H. K. Weir, J. M. Elwood, H. Tsukuma, S. Koifman, G. A. e Silva, S. Francisci, M. Santaquilani, A. Verdecchia, H. H. Storm and J. L. Young (2008). "Cancer survival in five continents: a worldwide population-based study (CONCORD)." The Lancet Oncology **9**(8): 730-756.

Datta, S., R. Budhauliya, B. Das, S. Chatterjee, Vanlalhmuaka and V. Veer (2015). "Next-generation sequencing in clinical virology: Discovery of new viruses." World J Virol **4**(3): 265-276.

De Paoli, P. and A. Carbone (2013). "Carcinogenic viruses and solid cancers without sufficient evidence of causal association." Int J Cancer **133**(7): 1517-1529.

Dela Cruz, F. N., F. Giannitti, L. Li, L. W. Woods, L. Del Valle, E. Delwart and P. A. Pesavento (2013). "Novel Polyomavirus associated with Brain Tumors in Free-Ranging Raccoons, Western United States." Emerging Infectious Diseases **19**(1): 77-84.

Eilebrecht, S., A. Hotz-Wagenblatt, V. Sarachaga, A. Burk, K. Falida, D. Chakraborty, E. Nikitina, C. Tessmer, C. Whitley, C. Sauerland, K. Gunst, I. Grewe and T. Bund (2018). "Expression and replication of virus-like circular DNA in human cells." Sci Rep **8**(1): 2851.

Epstein, M. M., E. T. Chang, Y. Zhang, T. T. Fung, J. L. Batista, R. F. Ambinder, T. Zheng, N. E. Mueller and B. M. Birmann (2015). "Dietary pattern and risk of hodgkin lymphoma in a population-based case-control study." Am J Epidemiol **182**(5): 405-416.

Falida, K., S. Eilebrecht, K. Gunst, H. zur Hausen and E. M. de Villiers (2017). "Isolation of Two Virus-Like Circular DNAs from Commercially Available Milk Samples." Genome Announc **5**(17).

Farlow, D. W., X. Xu and T. D. Veenstra (2009). "Quantitative measurement of endogenous estrogen metabolites, risk-factors for development of breast cancer, in commercial milk products by LC-MS/MS." J Chromatogr B Analyt Technol Biomed Life Sci **877**(13): 1327-1334.

Farvid, M. S., E. Cho, W. Y. Chen, A. H. Eliassen and W. C. Willett (2015). "Adolescent meat intake and breast cancer risk." Int J Cancer **136**(8): 1909-1920.

Fernandez, C., R. Nunez-Ramirez, M. Jimenez, G. Rivas and R. Giraldo (2016). "RepA-WH1, the agent of an amyloid proteinopathy in bacteria, builds oligomeric pores through lipid vesicles." Sci Rep **6**: 23144.

Funk, M., K. Gunst, V. Lucansky, H. Muller, H. Zur Hausen and E. M. de Villiers (2014). "Isolation of protein-associated circular DNA from healthy cattle serum." Genome Announc **2**(4).

Gadducci, A., N. Biglia, P. Sismondi and A. R. Genazzani (2005). "Breast cancer and sex steroids: Critical review of epidemiological, experimental and clinical investigations on etiopathogenesis, chemoprevention and endocrine treatment of breast cancer." Gynecological Endocrinology **20**(6): 343-360.

Gannon, O. M., A. Antonsson, I. C. Bennett and N. A. Saunders (2018). "Viral infections and breast cancer - A current perspective." Cancer Lett **420**: 182-189.

Ghoussaini, M., P. D. P. Pharoah and D. F. Easton (2013). "Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning?" Am J Pathol **183**(4): 1038-1051.

Glenn, W. K., B. Heng, W. Delprado, B. Iacopetta, N. J. Whitaker and J. S. Lawson (2012). "Epstein-Barr virus, human papillomavirus and mouse mammary tumour virus as multiple viruses in breast cancer." PLoS One **7**(11): e48788.

Golaszewska, E., E. Kurowska, M. Duk and J. Koscielak (2003). "Paul-Bunnell antigen and a possible mechanism of formation of heterophile antibodies in patients with infectious mononucleosis." Acta Biochim Pol **50**(4): 1205-1211.

Gunst, K., H. Zur Hausen and E. M. de Villiers (2014). "Isolation of bacterial plasmid-related replication-associated circular DNA from a serum sample of a multiple sclerosis patient." Genome Announc **2**(4).

Haahr, S., N. Koch-Henriksen, A. Moller-Larsen, L. S. Eriksen and H. M. Andersen (1995). "Increased risk of multiple sclerosis after late Epstein-Barr virus infection: a historical prospective study." Mult Scler **1**(2): 73-77.

Handelsman, J. (2004). "Metagenomics: application of genomics to uncultured microorganisms." Microbiol Mol Biol Rev **68**(4): 669-685.

Hanley-Bowdoin, L., E. R. Bejarano, D. Robertson and S. Mansoor (2013). "Geminiviruses: masters at redirecting and reprogramming plant processes." Nature Reviews Microbiology **11**: 777.

Huang, C., Y. Xie, L. Zhao, H. Ren and Z. Li (2013). "A naturally occurring defective DNA satellite associated with a monopartite begomovirus: evidence for recombination between alphasatellite and betasatellite." Viruses **5**(9): 2116-2128.

Jelcic, I., A. Hotz-Wagenblatt, A. Hunziker, H. Zur Hausen and E. M. de Villiers (2004). "Isolation of multiple TT virus genotypes from spleen biopsy tissue from a Hodgkin's disease patient: genome reorganization and diversity in the hypervariable region." J Virol **78**(14): 7498-7507.

Ji, J., J. Sundquist and K. Sundquist (2015). "Lactose intolerance and risk of lung, breast and ovarian cancers: aetiological clues from a population-based study in Sweden." Br J Cancer **112**(1): 149-152.

Jorgensen, T. S., Z. Xu, M. A. Hansen, S. J. Sorensen and L. H. Hansen (2014). "Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome." PLoS One **9**(2): e87924.
Joshi, D. and G. C. Buehring (2012). "Are viruses associated with human breast cancer? Scrutinizing the molecular evidence." Breast Cancer Res Treat **135**(1): 1-15.

Kabat, G. C., A. J. Cross, Y. Park, A. Schatzkin, A. R. Hollenbeck, T. E. Rohan and R. Sinha (2009). "Meat intake and meat preparation in relation to risk of postmenopausal breast cancer in the NIH-AARP diet and health study." Int J Cancer **124**(10): 2430-2435.

Kelsey, J. L., M. D. Gammon and E. M. John (1993). "Reproductive factors and breast cancer." Epidemiol Rev **15**(1): 36-47.

Khuder, S. A., A. B. Mutgi, E. A. Schaub and B. D. K. Tano (1999). "Meta-analysis of Hodgkin's disease among farmers." Scandinavian Journal of Work, Environment & Health **25**(5): 436-441.

Kuppers, R., K. Rajewsky, M. Zhao, G. Simons, R. Laumann, R. Fischer and M. L. Hansmann (1994). "Hodgkin disease: Hodgkin and Reed-Sternberg cells picked from histological sections show clonal immunoglobulin gene rearrangements and appear to be derived from B cells at various stages of development." Proc Natl Acad Sci U S A **91**(23): 10962-10966.

Lamberto, I., K. Gunst, H. Muller, H. Zur Hausen and E. M. de Villiers (2014). "Mycovirus-like DNA virus sequences from cattle serum and human brain and serum samples from multiple sclerosis patients." Genome Announc **2**(4).

Lawson, J. S., W. K. Glenn and N. J. Whitaker (2016). "Human Papilloma Viruses and Breast Cancer - Assessment of Causality." Front Oncol **6**: 207.

Lawson, J. S. and B. Heng (2010). "Viruses and Breast Cancer." Cancers **2**(2): 752.

Lawson, J. S., B. Salmons and W. K. Glenn (2018). "Oncogenic Viruses and Breast Cancer: Mouse Mammary Tumor Virus (MMTV), Bovine Leukemia

Virus (BLV), Human Papilloma Virus (HPV), and Epstein-Barr Virus (EBV)." Front Oncol **8**: 1.

Linos, E., W. C. Willett, E. Cho, G. Colditz and L. A. Frazier (2008). "Red meat consumption during adolescence among premenopausal women and risk of breast cancer." Cancer Epidemiol Biomarkers Prev **17**(8): 2146-2151.

Longkumer, T., S. Kamireddy, V. R. Muthyala, S. Akbarpasha, G. K. Pitchika, G. Kodetham, M. Ayaluru and D. Siddavattam (2013). "Acinetobacter phage genome is similar to Sphinx 2.36, the circular DNA copurified with TSE infected particles." Sci Rep **3**: 2240.

Male, M. F., S. Kraberger, D. Stainton, V. Kami and A. Varsani (2016). "Cycloviruses, gemycircularviruses and other novel replication-associated protein encoding circular viruses in Pacific flying fox (Pteropus tonganus) faeces." Infection, Genetics and Evolution **39**: 279-292.

Manuelidis, L. (2011). "Nuclease resistant circular DNAs copurify with infectivity in scrapie and CJD." J Neurovirol **17**(2): 131-145.

Marianne, E., D. S. W., A. Hans-Olov, K. Gunnar, L. Eiliv, M. Olav, M. Anders, S. Irma and T. Hrafn (1990). "Age at first birth, parity and risk of breast cancer: A meta-analysis of 8 studies from the nordic countries." International Journal of Cancer **46**(4): 597-603.

Maruyama, K., T. Oshima and K. Ohyama (2010). "Exposure to exogenous estrogen through intake of commercial milk produced from pregnant cows." Pediatr Int **52**(1): 33-38.

Mokili, J. L., F. Rohwer and B. E. Dutilh (2012). "Metagenomics and future perspectives in virus discovery." Current Opinion in Virology **2**(1): 63-77.

Moorman, P. G. and P. D. Terry (2004). "Consumption of dairy products and the risk of breast cancer: a review of the literature." Am J Clin Nutr **80**(1): 5-14.

Nartey, T., C. M. Mazzanti, S. Melana, W. K. Glenn, G. Bevilacqua, J. F. Holland, N. J. Whitaker, J. S. Lawson and B. G. Pogo (2017). "Mouse mammary tumor-like virus (MMTV) is present in human breast tissue before development of virally associated breast cancer." Infect Agent Cancer **12**: 1.

National Breast and Ovarian Cancer Center (2009). "Breast cancer risk factors: a review of the evidence." National Breast and Ovarian Cancer Center, Surry Hills, NSW.

Niel, C., L. Diniz-Mendes and S. Devalle (2005). "Rolling-circle amplification of Torque teno virus (TTV) complete genomes from human and swine sera and identification of a novel swine TTV genogroup." Journal of General Virology **86**(5): 1343-1347.

Peretti, A., P. C. FitzGerald, V. Bliskovsky, C. B. Buck and D. V. Pastrana (2015). "Hamburger polyomaviruses." Journal of General Virology **96**(4): 833-839.

Rajewska, M., K. Wegrzyn and I. Konieczny (2012). "AT-rich region and repeated sequences - the essential elements of replication origins of bacterial replicons." FEMS Microbiol Rev **36**(2): 408-434.

Rizvi, I., N. R. Choudhury and N. Tuteja (2015). "Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication." Archives of Virology **160**(2): 375-387.

Rosario, K., S. Duffy and M. Breitbart (2012). "A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics." Arch Virol **157**(10): 1851-1871.

Rosario, K., R. O. Schenck, R. C. Harbeitner, S. N. Lawler and M. Breitbart (2015). "Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins." Frontiers in Microbiology **6**(696).

Ruiz-Maso, J. A., N. C. Macho, L. Bordanaba-Ruiseco, M. Espinosa, M. Coll and G. Del Solar (2015). "Plasmid Rolling-Circle Replication." Microbiol Spectr **3**(1): Plas-0035-2014.

Samraj, A. N., H. Laubli, N. Varki and A. Varki (2014). "Involvement of a non-human sialic Acid in human cancer." Front Oncol **4**: 33.

Schowalter, R. M., D. V. Pastrana, K. A. Pumphrey, A. L. Moyer and C. B. Buck (2010). "Merkel Cell Polyomavirus and Two Previously Unknown Polyomaviruses Are Chronically Shed from Human Skin." Cell Host & Microbe **7**(6): 509-515.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nat Biotechnol **26**(10): 1135-1145.

Siegel, Kimberly and Jemal (2017). "Cancer statistics, 2017." CA: A Cancer Journal for Clinicians **67**(1): 7-30.

Simmonds, P., M. J. Adams, M. Benko, M. Breitbart, J. R. Brister, E. B. Carstens, A. J. Davison, E. Delwart, A. E. Gorbalenya, B. Harrach, R. Hull, A. M. King, E. V. Koonin, M. Krupovic, J. H. Kuhn, E. J. Lefkowitz, M. L. Nibert, R. Orton, M. J. Roossinck, S. Sabanadzovic, M. B. Sullivan, C. A. Suttle, R. B. Tesh, R. A. van der Vlugt, A. Varsani and F. M. Zerbini (2017). "Consensus statement: Virus taxonomy in the age of metagenomics." Nat Rev Microbiol **15**(3): 161-168.

Simone, D.-D., C. S. T., H. Dani, V.-Z. Arabel, B. J. A. and B. Gero (2009). "Cell type specific applicability of 5-ethynyl-2'-deoxyuridine (EdU) for dynamic proliferation assessment in flow cytometry." Cytometry Part A **75A**(6): 535-546.

Tang, P. and C. Chiu (2010). "Metagenomics for the discovery of novel human viruses." Future Microbiology **5**(2): 177-189.

Taylor, E. F., V. J. Burley, D. C. Greenwood and J. E. Cade (2007). "Meat consumption and risk of breast cancer in the UK Women's Cohort Study." Br J Cancer **96**(7): 1139-1146.

Taylor, R. E., C. J. Gregg, V. Padler-Karavani, D. Ghaderi, H. Yu, S. Huang, R. U. Sorensen, X. Chen, J. Inostroza, V. Nizet and A. Varki (2010). "Novel

mechanism for the generation of human xeno-autoantibodies against the nonhuman sialic acid N-glycolylneuraminic acid." J Exp Med **207**(8): 1637-1646.

Thacker, E. L., F. Mirzaei and A. Ascherio (2006). "Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis." Ann Neurol **59**(3): 499-503.

Thomas, D. B. (1993). "Breast cancer in men." Epidemiol Rev **15**(1): 220-231.

Towner, K. J. (2009). "Acinetobacter: an old friend, but a new enemy." Journal of Hospital Infection **73**(4): 355-363.

van der Meijden, E., R. W. A. Janssens, C. Lauber, J. N. Bouwes Bavinck, A. E. Gorbalenya and M. C. W. Feltkamp (2010). "Discovery of a New Human Polyomavirus Associated with Trichodysplasia Spinulosa in an Immunocompromized Patient." PLOS Pathogens **6**(7): e1001024.

Vlazny, D. A. and N. Frenkel (1981). "Replication of herpes simplex virus DNA: localization of replication recognition signals within defective virus genomes." Proc Natl Acad Sci U S A **78**(2): 742-746.

Vlazny, D. A. and R. W. Hyman (1985). "Errant processing and structural alterations of genomes present in a varicella-zoster virus vaccine." J Virol **56**(1): 92-101.

White, M. K., J. S. Pagano and K. Khalili (2014). "Viruses and human cancers: a long road of discovery of molecular paradigms." Clin Microbiol Rev **27**(3): 463-481.

Whitley, C., K. Gunst, H. Muller, M. Funk, H. Zur Hausen and E. M. de Villiers (2014). "Novel replication-competent circular DNA molecules from healthy cattle serum and milk and multiple sclerosis-affected human brain tissue." Genome Announc **2**(4).

Yeh, Y.-H., V. Gunasekharan and L. Manuelidis (2017). "A prokaryotic viral sequence is expressed and conserved in mammalian brain." Proceedings of the National Academy of Sciences of the United States of America **114**(27): 7118-7123.

zur Hausen, H., T. Bund and E. M. de Villiers (2017). "Infectious Agents in Bovine Red Meat and Milk and Their Potential Role in Cancer and Other Chronic Diseases." Curr Top Microbiol Immunol **407**: 83-116.

zur Hausen, H. and E. M. de Villiers (2014). "Cancer "causation" by infections--individual contributions and synergistic networks." Semin Oncol **41**(6): 860-875.

zur Hausen, H. and E. M. de Villiers (2015). "Dairy cattle serum and milk factors contributing to the risk of colon and breast cancers." Int J Cancer **137**(4): 959-967.

Appendix

A. List of abbreviations

BSA	bovine serum albumin
^{o}C	degree Celsius
DMSO	dimethylsulfoxid
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
E.coli	*Escherichia coli*
et al.	and others (et alii)
EtOH	ethanol
FCS	fetal calf serum
FFPE	formalin-fixed paraffin-embedded
Fig.	Figure
g	gram
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilo bases
LB	Luria-Bertani
m	milli
M	molar
µ	micro
min	minute
n	nano
NCBI	National Centre of Biotechnology
NEB	New England Biolabs
NGS	Next Generation Sequencing
PAGE	polyacrylamide gel-electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	preponderance of hydrogen ions
phi29Pol	29 polymerase
PNK	polynucleotide kinase
qPCR	quantitative polymerase chain reaction

RCA	rolling circle amplification
RNA	ribonucleic acid
rpm	rounds per minute
SDS	sodium dodecyl sulfate
sec	seconds
Taq	*Thermus aquaticus*
TBE	Tris/Borate/EDTA
Tris	Tris-(hydroxymethyl)-aminoethan
UV	ultraviolet
V	volt
v / v	volume per volume
w / v	weight per volume
www	world wide web
X-Gal	5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid

B. Laboratory equipment

Equipment	Model	Supplier
Centrifuge	5415D	Eppendorf
Vortex	Genie2™	Bender & Hobein AG
Rotator		
Lyophilization machine		
Mixer	5432	Eppendorf
Centrifuge	Sigma 2k15	SIGMA
Water bath	Julabo MP	JULABO
Electrophoresis power supply	GibcoBRL PS 305	Life Technologies
Thermocycler	FlexCycler^2	Analytik Jena
Centrifuge	Megafuge 1.0	Heraeus instruments
Centrifuge	Megafuge 1.0R	Heraeus instruments
Microprocessor pH meter	pH 537	WTW
Magnetic stirrer	REO Drehzahl-electronic	IKA IKAMAG
Shaker	Rocker25	Labnet
Shaking water bath	Type 3047	Kottermann
Laminar flow hood	SAFE2020	Thermo Scientific
CO_2 incubator	C200	Labotect
Blot imager	Chemidoc Touch Imaging System	Bio-Rad
Multimode plate reader	EnSpire	PerkinElmer
Real-Time PCR system	QuantStudio 5	Applied Biosystems
Gel documentation system	Molecular Imager Gel Doc XR	Bio-Rad

UV light plate	N90 LW 366 nm	Konrad Benda
Incubation shaker	Multitron Pro	INFORS HT
Material test chamber	Material test chamber FP 53	BINDER
Centrifuge	RC-5C	Sorvall
Transfer system	Trans-Blot Turbo	Bio-Rad
Spectrophotometer	Nanodrop2000	Thermo Scientific
Zeiss cell observer		

C. List of dairy products used for DNA isolation

Type of product	Product name	Name in the lab
Yoghurt	Bombay Lassi Mango - Giacomo	mi5
	Gut and Günstig - EDEKA	mi6
Milk	Gut and Günstig, Vollmilch, DE HE013 Eg	mi7
	Unsere Heinat, Bio Vollmilch, demeter, Heumilch, Milch ab BW, Wongen	mi8
	Schwälbehen, Bad Schwalbah, Vollmilch	mi9
	Schwartwaldmilch, Bio Vollmilch, Fräburg	mi10
	Berchtesgardener Land, Frishe Bergbavern, Milch	mi15
	Weihen stephan, Frische Milch 3.5%, Freising	mi16
Crème fraîche	Gut and günstig, Crème fraiche, EDEKA	mi11
	Demeter, Bio - crème fraiche, Berchtesgardener Land	mi12
Quark	Speisequark, 40%, Scwälbchen	mi13
	QUARK, 40%, Ehrmann	mi14
Butter	Irische Butter, Schüttent Lemmerholtz, 414b8 Neuss	mi15
	Weihenstephan, Butter	mi16

D. Supplemented figures and tables

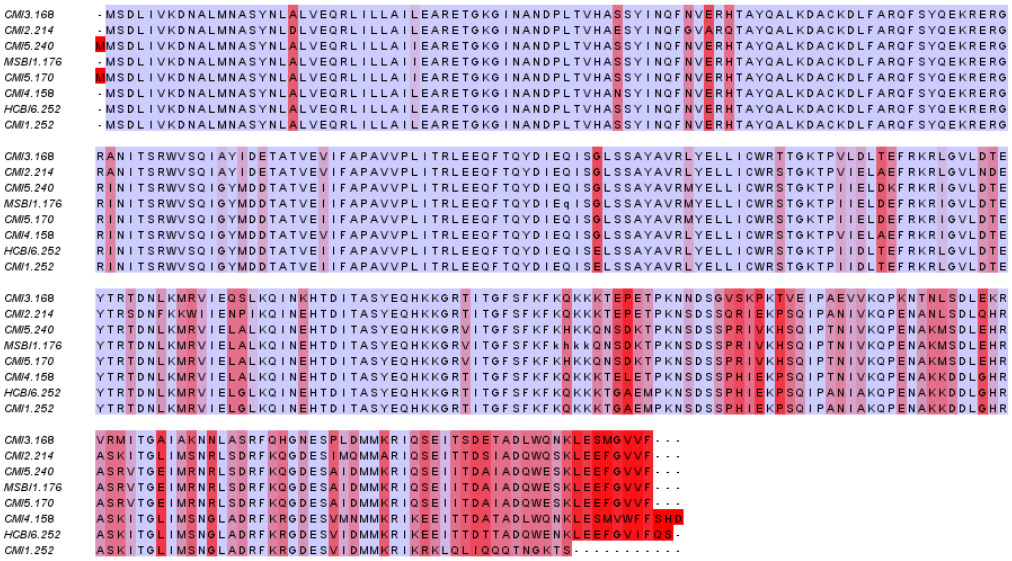


Figure S1 CLUSTAL alignment of putative replication proteins of BMMF1 including the newly isolated CMI5.170 and CMI5.240.

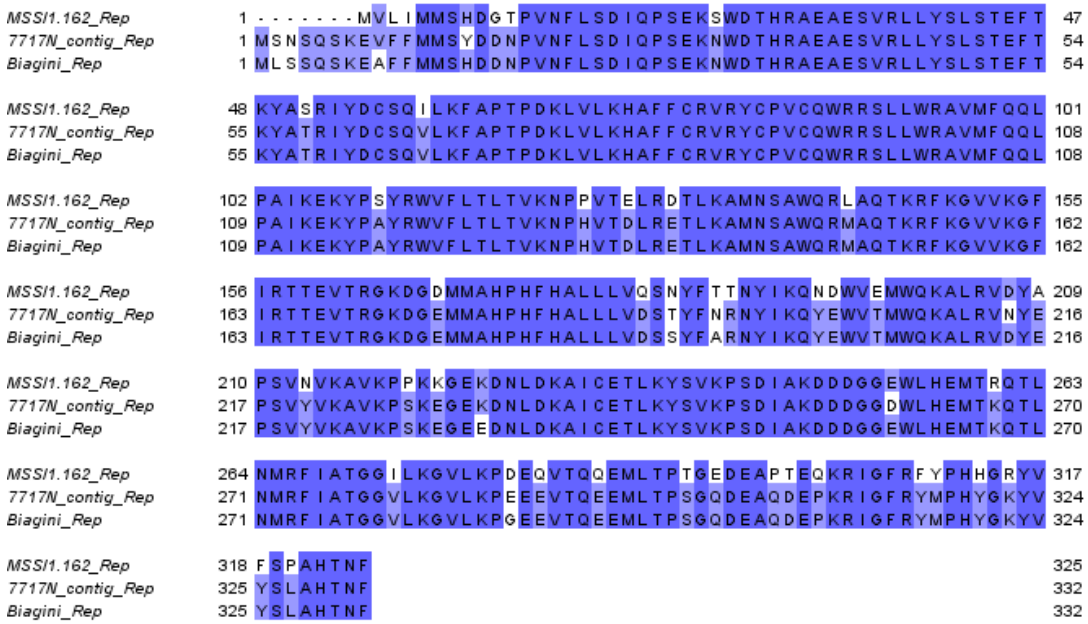


Figure S2 CLUSTAL alignment of putative replication proteins of MSSI1.162, 7717N_contig sequence isolated from breast tissue and the plasmid isolated from Biagini (KX838914).

Table S1 Information on breast cancer serum patients.

Pat. ID	Age	Right breast	Left breast
MaCa 001	53	Invasiv-lobuläres Mammakarzinom	
MaCa 002	69	bifokales Invasiv-duktales Mammakarzinom	Invasiv-duktales Mammakarzinom
MaCa 003	67	Invasiv-duktales Mammakarzinom	
MaCa 004	55	Invasiv-duktales Mammakarzinom rechts	
MaCa 005	57		Multifokales Invasiv-lobuläres Mammakarzinom
MaCa 006	71	Invasiv-duktales Mammakarzinom Rezidiv rechts	
MaCa 007	73		Invasives Mammakarzinom, spezieller Typ links
MaCa 008	49		multifokales Invasiv-lobuläres Mammakarzinom links
MaCa 009	68		Invasiv-lobuläres Mammakarzinom Zweitkarzinom links
MaCa 010	32		Invasives Mammakarzinom links, NST, G3, pN0
MaCa 011	30	medullärem Mammakarzinom	
MaCa 012	43	Invasiv-duktales Mammakarzinom rechts	
MaCa 013	45	Tubuläres Mammakarzinom	
MaCa 014	50	Muzinöses Mammakarzinom rechts Rezidiv	
MaCa 015	81		invasives Mammakarzinom NST links
MaCa 016	77	Invasiv-duktales Mammakarzinom rechts	
MaCa 017	77	Invasiv-duktales Mammakarzinom rechts	
MaCa 018	52	Invasiv-duktales Mammakarzinom rechts	
MaCa 019	60	Invasiv-duktales Mammakarzinom rechts	
MaCa 020	56		Invasives Mammakarzinom links
MaCa 021	50		Invasiv-duktales Mammakarzinom links
MaCa 022	52		Invasiv-duktales Mammakarzinom links
MaCa 023	54		Muzinöses Mammakarzinom links
MaCa 024	85		Invasiv-lobuläres Mammakarzinom links
MaCa 025	54		Primär ossär metastasiertes invasiv-duktales Mammakarzinom links
MaCa 026	50		bifokales Invasiv-lobuläres Mammakarzinom links
MaCa 027	48	Invasiv-lobuläres Mammakarzinom rechts	
MaCa 028	28	Invasiv-duktales Mammakarzinom rechts	
MaCa 029	64	Invasiv-duktales Mammakarzinom rechts	
MaCa 030	59		Invasiv-duktales Mammakarzinom links

Table S2 Age and sex information of the plasma samples tested.

Number	Age	Sex
1	21	♀
2	27	♂
3	22	♀
4	23	♂
5	22	♂
6	26	♂
7	23	♂
8	30	♂
9	25	♂
10	30	♂
11	23	♀
12	21	♀
13	32	♀
14	22	♂
15	22	♂
16	23	♀
17	63	♂

Number	Age	Sex
18	22	♂
19	20	♀
20	27	♂
21	26	♀
22	26	♂
23	20	♂
24	29	♂
25	23	♂
26	27	♂
27	20	♂
28	40	♀
29	20	♀
30	56	♀
31	22	♀
32	24	♂
33	58	♂
34	23	♀

Table S3 Information about reads obtained after sequencing and the corresponding barcode of each library sample.

SAMPLE_ID	READ_COUNT	BARCODE		SAMPLE_ID	READ_COUNT	BARCODE
7714T	71612013	ATCACG		7714N	22717321	GATCAG
7714T	71612013	ATCACG		7714N	22717321	GATCAG
7715T	19112680	CGATGT		7715N	19255549	TAGCTT
7715T	19112680	CGATGT		7715N	19255549	TAGCTT
7716T	26206563	TTAGGC		7716N	17852234	GGCTAC
7716T	26206563	TTAGGC		7716N	17852234	GGCTAC
7717T	21383940	TGACCA		7717N	30507604	GTGAAA
7717T	21383940	TGACCA		7717N	30507604	GTGAAA
7718T	25962669	ACAGTG		7718N	26082825	ACTGAT
7718T	25962669	ACAGTG		7718N	26082825	ACTGAT
7722T	30303146	CAGATC		7722N	28414396	GCCAAT
7722T	30303146	CAGATC		7722N	28414396	GCCAAT
7724T	22709478	ACTTGA		7724N	27526352	CTTGTA
7724T	22709478	ACTTGA		7724N	27526352	CTTGTA

E. List of figures

Figure 1.1 Global estimated incidence of colon (A) and breast cancer (B), GLOBOCAN 2012 (IARC). High risk regions for cancer incidence are shown in dark blue color and show similarity for the two types of cancer and the same applies for low risk counties indicated with light blue color.

Figure 1.2 Schematic view of the novel episomal circular DNAs isolated from milk, bovine serum and tissue and blood samples from MS patients (adapted from zur Hausen, Bund and de Villiers 2017).

Figure 1.3 Genetic diversity of CRESS-DNA viruses. Comparison of the Rep sequences of more than 600 CRESS-DNAs with 10 representative Rep sequences from viruses classified in the families *Gemonoviridae, Geminiviridae*, *Nanoviridae* and *Circiviridae,* and a group of alha satellites that are associated with geminiviruses

Figure 3.7 RT-qPCR data representing the enrichment of Nanoluc plasmid DNA in each extraction method before and after RCA as well as after RCA upon Plasmid-Safe DNase digestion. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

Figure 3.8 A) Circular representation of the contig found in breast NAT tissue sample. **B)** Alignment of the circular contig with the previously isolated MSSI1.162 and the plasmid gene isolated from Biagini (KX838914).

Figure 3.9 PCR with specific back-to-back primers on the seven breast tumor tissue samples (7714T-77124T) and the corresponding NAT controls (7714N-7724N) confirming the presence of the new isolate only in the expected (7717N) sample with approximately 1600 bp size.

Figure 3.10 A) PCR after DpnI digestion on DNA extracted from L-1236 cells after transfection with MSBI1.176 using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction and DpnI-digested circular MSBI1.176 was used as a template for a negative control reaction. **B)** qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-1236 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA. **C)** qPCR data representing the ratio of MSBI1.176 DpnI digested DNA to MSBI1.176 non-DpnI digested DNA at different harvesting time points of L-1236 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

Figure 3.11 A) PCR after DpnI digestion on DNA extracted from L-428 cells after transfection with MSBI1.176 using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction and DpnI-digested circular MSBI1.176 was used as a template for a negative control reaction. **B)** RT-qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-428 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the

amount of input DNA. **C)** RT-qPCR data representing the ratio of MSBI1.176 DpnI digested DNA to MSBI1.176 non-DpnI digested DNA at different harvesting time points of L-428 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

Figure 3.12 RT-qPCR data representing the number of CMI1.252 molecules per cell at different harvesting time points of L-1236 cells. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

Figure 3.13 PCR after DpnI digestion on DNA extracted from L-428 cells after transfection with sg613 using sg613-specific back-to-back primers. Circular MSBI1.176 and sg613 were used as template for a positive control reaction and DpnI-digested circular MSBI1.176 and sg613 were used as template for a negative control reaction.

Figure 3.14 A) PCR after DpnI digestion on DNA extracted from L-428 cells after transfection with MSBI1.176 and co-transfection with MSBI1.176 and sg613 using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 and sg613 were used as template for a positive control reaction and DpnI-digested circular MSBI1.176 and sg613 were used as template for a negative control reaction. **B)** RT-qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-428 cells for the single MSBI1.176 condition and the co-transfection with sg613. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA. **C)** RT-qPCR data representing the ratio of MSBI1.176 DpnI digested DNA to MSBI1.176 non-DpnI digested DNA at different harvesting time points of L-428 cells for the single condition of MSBI1.176 and the co-transfection with sg613. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

Figure 3.15 A) PCR after DpnI digestion on DNA extracted from L-1236 cells after transfection with MSBI1.176 and MSBI1.176 with external addition of 100 μM Neu5Gc using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction and DpnI-digested circular MSBI1.176 was used as a template for a negative control reaction. **B)** qPCR data representing the number of MSBI1.176 molecules per cell at different harvesting time points of L-1236 cells with and without addition of external Neu5Gc. A region within the beta actin gene was amplified in parallel for each sample using specific primers, in order to control for the amount of input DNA.

Figure 3.16 PCR analysis of DNA extracted from supernatants collected after transfection of L-1236 cells with MSBI1.176 and MSBI1.176 pre-treated with Neu5Gc after day 3 post transfection and at each harvesting time point using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction. Transfection with pUC19 vector and 1xTE buffer were used as negative control reactions.

Figure 3.17 PCR on DNA extracted from L-1236 cells after infection with supernatants from transfection with MSBI1.176 and controls using MSBI1.176-specific back-to-back primers. Circular MSBI1.176 was used as a template for a positive control reaction. The red color letters indicate the conditions with 100-fold excess of Neu5Gc and the green color the ones without addition of external Neu5Gc.

Figure 3.18 Luminescent signal after transfection of L-428 cells with pcDNA3.1(-) control vector, MSBI1.176 Rep WT OE, MSBI1.176 Rep MUT OE, MSBI1.176 linear and MSBI1.176 circular genome measured at three different time points post transfection.

Figure 3.19 Summary of cell proliferation assay in L-428 cells after transfection with pcDNA3.1(-) control vector, MSBI1.176 Rep WT OE protein and MSBI1.176 Rep Mut protein. **A)** representative EdU positive cells at 48h post transfection, **B)** percentage of EdU positive cells after 24h, 48h and 72h for all three conditions, C) western blot analysis suing anti-His Rep antibody and alpha tubulin as loading control. No difference was observed for the two genomes.

Figure 3.20 Box plot analysis based on ELISA results. Comparison between 50 Hodgkin's Lymphoma serum samples, 30 breast cancer serum samples and 30 plasma controls using Wild type (WT) Rep MSBI1.176 protein as antigen.

F. List of tables

G. Publications

Eilebrecht, S., Hotz-Wagenblatt, A., Sarachaga, V., Burk, A., Falida, K., Chakraborty, D., Nikitina, E.,Tessmer, C., Whitley, C., Sauerland, C., Gunst, K., Grewe, I., Bund, T. (2018). Expression and replication of virus-like circular DNA in human cells. ***Sci Rep.*** **8** (1)

Falida, K., S. Eilebrecht, K. Gunst, H. zur Hausen and E. M. de Villiers (2017). Isolation of Two Virus-Like Circular DNAs from Commercially Available Milk Samples. ***Genome Announc.*** **5** *(17)*

ACKNOWLEDGEMENTS

This dissertation represents my personal work, but it would have not been possible without the support of friends, family and colleagues. First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. H. zur Hausen for the continuous support of my PhD study and related research, for sharing his immense knowledge, for inspiring me and being patient with me during the last three years. I am grateful for trusting me with such an interesting and exciting topic and making me part of your group at the DKFZ. Your guidance helped me in all time of experimental work and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

I would also like to thank Prof. E-M. de Villiers for teaching me how to analyze DNA sequences and for the time she spent analyzing the data, as well as for her questions and comments which incented me to widen my research from various perspectives. In addition, I thank the rest of my thesis committee: Prof. M. von Knebel-Döberitz and Prof. M. Heikenwälder, for their insightful comments and encouragement.

My sincere thanks also go to the two Post-Docs of our group, Dr. T. Bund (currently head of the group) and Dr. S. Eilebrecht, who were always willing to find time for teaching me, answering my questions and helping me solve my problems. Without their precious support it would not have been possible to conduct this research. Special thanks go to Agnes for the analysis of the Next Generation Sequencing data, to Victor for his help throughout my research, to Amelie for all the fun we had in the last three years, to Karin and Corinna who taught me most of the techniques and the rest of my fellow lab mates for their support and the nice time we had together.

I want to thank Prof. Dr. Tsivgoulis for his endless support over the last years. His words and advices were accompanying me through the PhD and helped me overcome difficult moments.

I am also thankful to Oryx for the financial support and to the Bayer Science and Education Foundation for offering me a fellowship in order to conduct my research.

Thanks to all my friends, Chistos, Christos, Fereydoon, Georgia, Rula, Lisa, for filling the last three years with beautiful moments and supporting me through my PhD.

Many thanks go to my wonderful family, my parents and brother, who are always there for me, encouraging my choices and helping me in every step of the way. Thank you for making me a better person and for teaching me to appreciate learning and knowledge.

Finally yet importantly, I thank Alexandros, for his understanding and patience, for holding my hand through this journey and for motivating me whenever I needed it the most.